TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Atanas Boev

Perceptually Optimized Visualization on Autostereoscopic 3D Displays

Julkaisu 1063 • Publication 1063

Tampere 2012

Tampereen teknillinen yliopisto. Julkaisu 1063
Tampere University of Technology. Publication 1063

Atanas Boev

Perceptually Optimized Visualization on Autostereoscopic 3D Displays

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium S4, at Tampere University of Technology, on the 31st of August 2012, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2012

ISBN 978-952-15-2892-7 (printed)
ISBN 978-952-15-3038-8 (PDF)
ISSN 1459-2045

Abstract

The family of displays, which aims to visualize a 3D scene with realistic depth, are known as “3D displays”. Due to technical limitations and design decisions, such displays create visible distortions, which are interpreted by the human vision as artefacts. In absence of visual reference (e.g. the original scene is not available for comparison) one can improve the perceived quality of the representations by making the distortions less visible. This thesis proposes a number of signal processing techniques for decreasing the visibility of artefacts on 3D displays.

The visual perception of depth is discussed, and the properties (*depth cues*) of a scene which the brain uses for assessing an image in 3D are identified. Following the physiology of vision, a taxonomy of 3D artefacts is proposed. The taxonomy classifies the artefacts based on their origin and on the way they are interpreted by the human visual system.

The principles of operation of the most popular types of 3D displays are explained. Based on the display operation principles, 3D displays are modelled as a signal processing channel. The model is used to explain the process of introducing distortions. It also allows one to identify which optical properties of a display are most relevant to the creation of artefacts. A set of optical properties for dual-view and multiview 3D displays are identified, and a methodology for measuring them is introduced. The measurement methodology allows one to derive the angular visibility and crosstalk of each display element without the need for precision measurement equipment. Based on the measurements, a methodology for creating a *quality profile* of 3D displays is proposed. The quality profile can be either simulated using the angular brightness function or directly measured from a series of photographs. A comparative study introducing the measurement results on the visual quality and position of the sweet-spots of eleven 3D displays of different types is presented. Knowing the sweet-spot position and the quality profile allows for easy comparison between 3D displays. The shape and size of the passband allows depth and textures of a 3D content to be optimized for a given 3D display.

Based on knowledge of 3D artefact visibility and an understanding of distortions introduced by 3D displays, a number of signal processing techniques for artefact mitigation are created. A methodology for creating anti-aliasing filters for 3D displays is proposed. For multiview displays, the methodology is extended towards so-called *passband optimization* which addresses Moiré, fixed-pattern-noise and ghosting artefacts, which are characteristic for such displays. Additionally, design of tuneable anti-aliasing filters is presented, along with a framework which allows the user to select the so-called *3d sharpness* parameter according to his or her preferences. Finally, a set of real-time algorithms for view-point-based optimization are presented. These algorithms require active user-tracking, which is implemented as a combination of face and eye-tracking. Once the observer position is known, the image on a stereoscopic display is optimised for the derived observation angle and distance. For multiview displays, the combination of precise light re-direction and less-precise face-tracking is used for extending the head parallax. For some user-tracking algorithms, implementation details are given, regarding execution of the algorithm on a mobile device or on desktop computer with graphical accelerator.

Preface

The research for this thesis has been conducted at Tampere University of Technology (TUT) during the years 2005-2012. I would like to express my gratitude to my supervisor Prof Karen Egiazarian for providing the great opportunity to work in the area of signal processing, and his guidance through the years of my work and study. I am especially grateful to my co-supervisor Dr Atanas Gotchev, for introducing me to the area of 3D display image processing, and for his continuous support and guidance during my time in Tampere as a visiting researcher and postgraduate student, trough 8 years, 32 publications, 18 project reports and multiple dissemination activities. I highly appreciate the feedback and insightful comments of Dr Phil Surman (De Montfort University, UK) and Dr Òscar Divorra (Telefónica I+D, Spain) who were pre-examiners of my thesis. I am indebted to Dr Martin Schrader (Nokia Research Centre, Finland) and Dr Phil Surman for agreeing to be opponents in the public defence of my thesis.

I am grateful to the co-authors of the papers for their help and collaboration. I would like to thank especially to Robert Bregovic for his valuable insights and help on building the visual optimization framework, which serves as “fundamental frequency” of this thesis. I am grateful to Satu Jumisko-Pyykkö, Dominik Strohmeier, Gozde Bozdagi Akar, Tolga Çapın, Mihail Georgiev, Kalle Raunio, Danilo Hollosi, Damyan Damyanov, as without their collaboration this thesis would not be possible. Special thanks to Dr Alessandro Foi for his help with modelling of the angular visibility and overall support on multiple research topics. A am also thankful to my co-authors Miska Hannuksela, Maija Mikkola, Timo Utriainen, Prof Jaakko Astola, Vladimir Katkovnik, Anil Aksay, Tapio Saramäki, Chavdar Kalchev, Ilian Todorov, Lina Jin, Marianne Hanhela, Antti Tikanmäki, Tomi Haustola for their valuable contribution to my research. I am grateful to Jyrki Häyrynen and Jarkko Pekkarinen for helping with the measurement experiments, as the results from these measurements are indispensable for the display modelling proposed in this thesis.

I would like to express my gratitude to Prof Jaakko Astola for inviting be as a visiting researcher to the Tampere International Centre for Signal Processing. I am indebted to Pirkko Ruotsalainen, Virve Larmila, Elina Orava and Ulla Siltaloppi for providing help on numerous occasions during my stay in the Department. I would also like to acknowledge my colleagues and friends for making my life at academia inspiring and enjoyable experience. I would like to thank wholeheartedly to Stanislav Stankovic, Annamaria Mesaros, Vladislav Uzunov, Kostadin Dabov, Susanna Minasyan, Jugoslava Asimovic, Nadezhda Gotcheva, Aram Danielyan, and Toni Heittola.

I am grateful to the Institute of Signal Processing (led by Prof Moncef Gabbouj) and later, Department of Signal Processing (led by Prof Ari Visa) for providing inspiring and friendly environment during the years of my post-graduate studies. I am grateful for receiving a grant from the Nokia Foundation, and travel support from the Tampere Doctoral Programme in Information Science and Engineering.

My warmest thanks belong to my family for their unconditional support and love. I thank with all my heart to my parents Zoia and Rumen, my sisters Anna and Maksima, my wife Silviya and my son Alexander.

Atanas Boev,
Tampere, 11.08.2012

Contents

List of publications

The thesis consists of a summary and the following original publications:

[P01] A. Boev, R. Bregovic and A. Gotchev, "Visual-quality evaluation methodology for multiview displays," *Displays,* vol. 33, no. 2, pp. 103-112, April 2012.

[P02] Atanas Gotchev, Gozde Bozdagi Akar, Tolga Capin, Dominik Strohmeier, Atanas Boev, "Three-Dimensional Media for Mobile Devices", *Proceedings of the IEEE*, April 2011, Vol. 99, 4, pp. 708-737, DOI: 10.1109/JPROC.2010.2103290

[P03] A. Boev, R. Bregovic, A. Gotchev, "Methodology for design of antialiasing filters for autostereoscopic displays", Special issue on *Advanced Techniques on Multirate Signal Processing for Digital Information Processing, Journal of IET Signal Processing,* Volume 5, Issue 3, June 2010, pp. 333-343

[P04] A. Boev, R. Bregovic, A. Gotchev, "Measuring and modeling per-element angular visibility in multiview displays", *Special issue on 3D displays, Journal of Society for Information Display,* Sept. 2010 Vol. 18, No. 09, pp. 686–697

[P05] A. Boev, A. Gotchev, "Comparative study of autostereoscopic displays for mobile devices", *Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V.* Edited by Akopian, David; Creutzburg, Reiner; Snoek, Cees G. M.; Sebe, Nicu; Kennedy, Lyndon. Proceedings of the SPIE, Volume 7881, pp. 78810B-78810B-12 (2011)

[P06] A. Boev, R. Bregovic, A. Gotchev, "Design of tuneable anti-aliasing filters for multiview displays", *Stereoscopic Displays and Applications XXII, Proc. SPIE 7863*, 78630F (2011), DOI: 10.1117/12.873465

[P07] A. Boev, R. Bregovic, D. Damyanov, A. Gotchev, "Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays", in Proc. of *The 2009 International Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2009*, Helsinki, Finland, 2009

[P08] A. Boev, D. Hollosi, Atanas Gotchev and Karen Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content", *Stereoscopic Displays and Applications XX, Proc. SPIE 7237*, 72371F (2009), DOI:10.1117/12.807185

[P09] A. Boev, K. Raunio, M. Georgiev, A. Gotchev, K. Egiazarian, "Opengl-Based Control of Semi-Active 3D Display," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008* , vol., no., pp.125-128, 28-30 May 2008 doi: 10.1109/3DTV.2008.4547824

[P10] A. Boev, K. Raunio, A. Gotchev and K. Egiazarian, *Stereoscopic Displays and Applications XIX* "GPU-based algorithms for optimized visualization and crosstalk mitigation on a multiview display", *Proc. SPIE 6803*, 68030K (2008), DOI:10.1117/12.761785

[P11] A. Boev, A. Gotchev, K. Egiazarian, "Crosstalk Measurement Methodology for Auto-Stereoscopic Screens," Proc. *3DTV Conference, 2007*, pp.1-4, 7-9 May 2007 doi: 10.1109/3DTV.2007.4379396

[P12] A. Boev, M. Georgiev, A. Gotchev, N. Daskalov and K. Egiazarian "Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display", in *Proc. 17th European Signal Conference EUSIPCO 2009*, Glasgow, Scotland, August 2009

List of acronyms

HVS ..... Human Visual System

LGN ..... Lateral Geniculate Nucleus

CSF ..... Contrast Sensitivity Function

SDPG ..... Stereoscopic Displays with Passive Glasses.

TFT ..... Thin Film Transistor

LCD ..... Liquid Crystal Display

IPD ..... Interpupillary Distance

OVD ..... Optimal Viewing Distance

FPS ..... Frames Per Second

EPI ..... Epipolar Plane Image

ARM ..... Advanced RISC Machine architecture

DSP ..... Digital Signal Processing / Digital Signal Processor

GPU ..... Graphical Processing Unit

1 Introduction

The purpose of a display is to convey a visual message. Naturally, one of the design goals in display development is to achieve more convincing and easier for the eye image output. Stereoscopic 3D displays are the current technological advance aiming at a better 3D scene representation.

A real three-dimensional scene radiates a complex waveform. However, not all properties of that waveform are visually significant. Some of them, such as light phase and polarization are not perceivable. Other scene information, for example wavelength and luminance' is crudely encoded by the visual system, which makes the exact replica of that property visually indistinguishable from its rough approximation.

For the purpose of simplified design, stereoscopic displays are meant to recreate only the visually important properties of the scene, while omitting the "unnecessary" ones. This

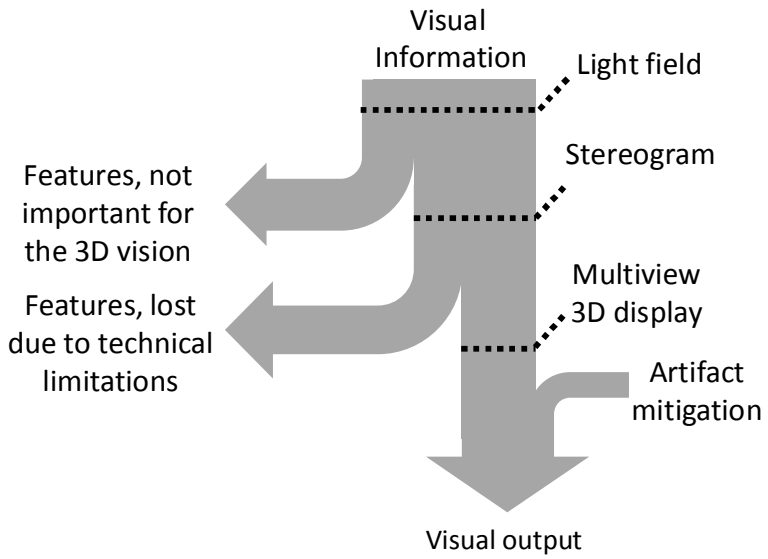


Figure 1. The process of artefact creation and mitigation.

decreases the amount of data which needs to be processed, while keeping the reduction unperceivable. However, due to technical limitations, additional scene information is lost. Since there is perceptual difference, the scene is seen as being unnatural and the differences are interpreted by the human visual system (HVS) as artefacts. Due to the way the HVS works, some structural distortions are more visible than others. Once lost, the missing visual information cannot be fully reconstructed. However, in a situation where the intended representation is not available as visual reference, some artefacts are more objectionable than others. If one can make artefacts less noticeable, the scene will be perceived as a more pleasant representation with higher subjective quality. In a nutshell, this dissertation discusses techniques to decrease the visibility of artefacts on a 3D display.

1.1 Objective

The process of artefact creation and mitigation can be regarded as having four conceptual stages as show in Figure 1. The overall visual information of a scene can be described using the amount of light in every direction and through every point the 3D volume encompassing the

scene. Such description is also known as the *light field* of a scene [1]. However, human vision utilizes only a fraction of this information. Some scene features are disregarded in the process of visual perception. If a visual representation of a scene contains only the information which matters for the HVS, the result would be visually indistinguishable from the light field the original scene. Such representation is sometimes referred to as a *stereogram* [2].

The ideal 3D display should be able to create a perfect stereogram of a scene. In reality, important scene features are misrepresented, which leads to perceivable distortions. Such distortions are recognized as artefacts. By introducing artefact mitigation techniques one can improve the subjectively perceived visual quality of the display. Visual optimization requires understanding of two factors – the important *visual properties* of a 3D scene and the relevant

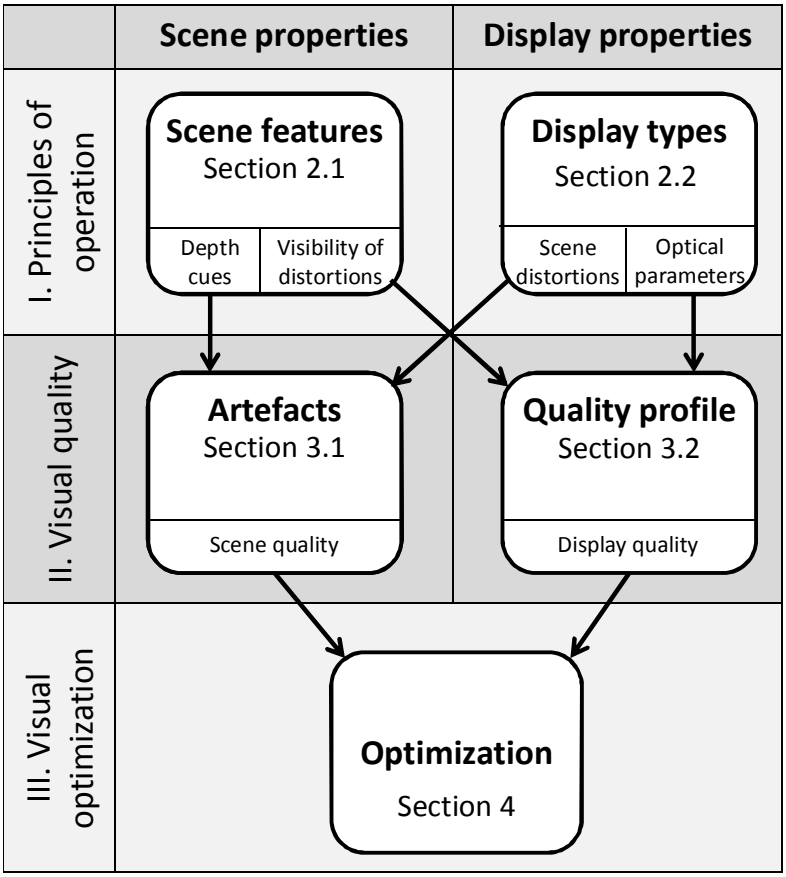


Figure 2. Outline of the approach.

display properties that allow the scene to be shown in 3D. Visual optimization of a 3D scene aims to maximize the "important" 3D features, while suppressing the "annoying" ones. The resulting scene is not true to the original. Instead, the aim is to achieve the most visually pleasant output possible for a given display.

1.2 Outline

The dissertation consists of two parts. The first part contains a summary, which introduces the problem, presents the state-of-the-art, and describes the scientific approach. The second part contains a list of original publications. The outline of the summary is given in Figure 2. Section 2.1 discusses which 3D features are visually important, and how these features can be included in a 3D scene representation. In Section 2.2, a 3D display classification is presented. The

classification is done based on the method each display uses to recreate the stereoscopic image. In Section 3.1, knowledge of display specifics and HVS properties is used to explain the appearance and visibility of artefacts on 3D displays. Section 3.2 discusses which optical properties of a 3D display are important from visual quality point of view, and presents a methodology to measure these properties. These measurements allow one to derive the so-called *quality profile* of a given 3D display. In Section 4, the understanding of artefact visibility and knowledge of optical quality is used for a set of signal processing algorithms which aim at a visual optimization of a 3D scene. Section 5 gives the conclusions.

An overview of how 3D artefacts affect perceptual quality is presented in [P02] and [P08]. In [P08], classification of artefacts and a framework for simulation is also presented. Measuring of crosstalk and angular visibility of a 3D display is started in [P11]. The measurement methodology is further developed in [P04] where it includes deriving the interleaving topology and independent per-pixel angular visibility without the need of precise camera positioning. Comparative study of the optical quality of 3D displays is done in [P02] and [P05]. In [P01], optical measurements are used to derive so-called *display passband*, which can be used as an indicator of the perceptual quality of a given 3D display. The design of anti-aliasing filters for 3D displays is discussed in [P07] and is extended towards simultaneous removal of moiré artefacts and mitigation of fixed-pattern noise in [P03]. Furthermore, design of tuneable and content-aware image filters where the user can select the preferred level of so-called *3d sharpness*, is described in [P06]. Additional visual optimization algorithms for 3D displays are presented in [P09] (extended head-parallax) and [P10] (crosstalk mitigation).

2 Principles of 3D visualization

2.1 3D scene characteristics

An example of a real 3D display is the window display shown in Figure 3. It allows the pedestrians to see a 3D scene from a wide range of angles and distances or take photos with arbitrary focal length. They can interact with the light of the scene by pointing a flashlight and making objects change colour or cast shadows. The store display emits complex light radiation with a wide spectrum, and the resolution of the scene is limited only by the size of its atoms.

The ideal 3D display would allow the store owner to replace the objects on display with a special window surface which contains the *perfect visual replica* of the scene. However, not all visual properties of the scene are equally important. The real-life scenario does not require the window to react to a flashlight or allow the observer to walk through the scene. The ultraviolet

Figure 3. Real 3D scene on display in a shop window

spectrum or light polarization is not seen by the naked eye, and thus, they do not need to be visualized. Removing this information produces a *redundancy-free replica* of the scene. In a typical use case, redundancy-free replica is also visually indistinguishable representation of the scene. Failure in creating redundancy-free and visually-indistinguishable replica leads to visible distortions. In order to avoid this, one needs to know which light properties are important, and which scene features are relevant for perceiving the scene in 3D.

2.1.1 Binocular vision

The study of the HVS could be separated into two parts - visual perception and visual cognition. The primary visual perception and the phenomena known as *early vision* are studied in anatomy [3] [4]. Visual cognition as a higher level processing function of the brain is a subject of psychology [3] [5].

Visual perception is result of a number of optical and neural transformations. The structural model of the HVS is shown in Figure 4a. First, the light passes through the eye's cornea and lens, by which it is focused in the retina. Notably, only 6% of earth species have such an eye structure, while more than 77% have compound eyes similar to those of a fly [5]. The retina is considered to be part of the brain, sequestered in early stages of embryonic development [6]. The image projected on the retina is transformed to neural impulses by two types of photoreceptors - cone and rod cells. The cone cells are of three types – L-cones, M-cones and S-cones. Each cone type can be thought (to a crude approximation) as sensitive to red, green and blue colour components of the projected light. In fact, each photoreceptor type reacts to a wide range of

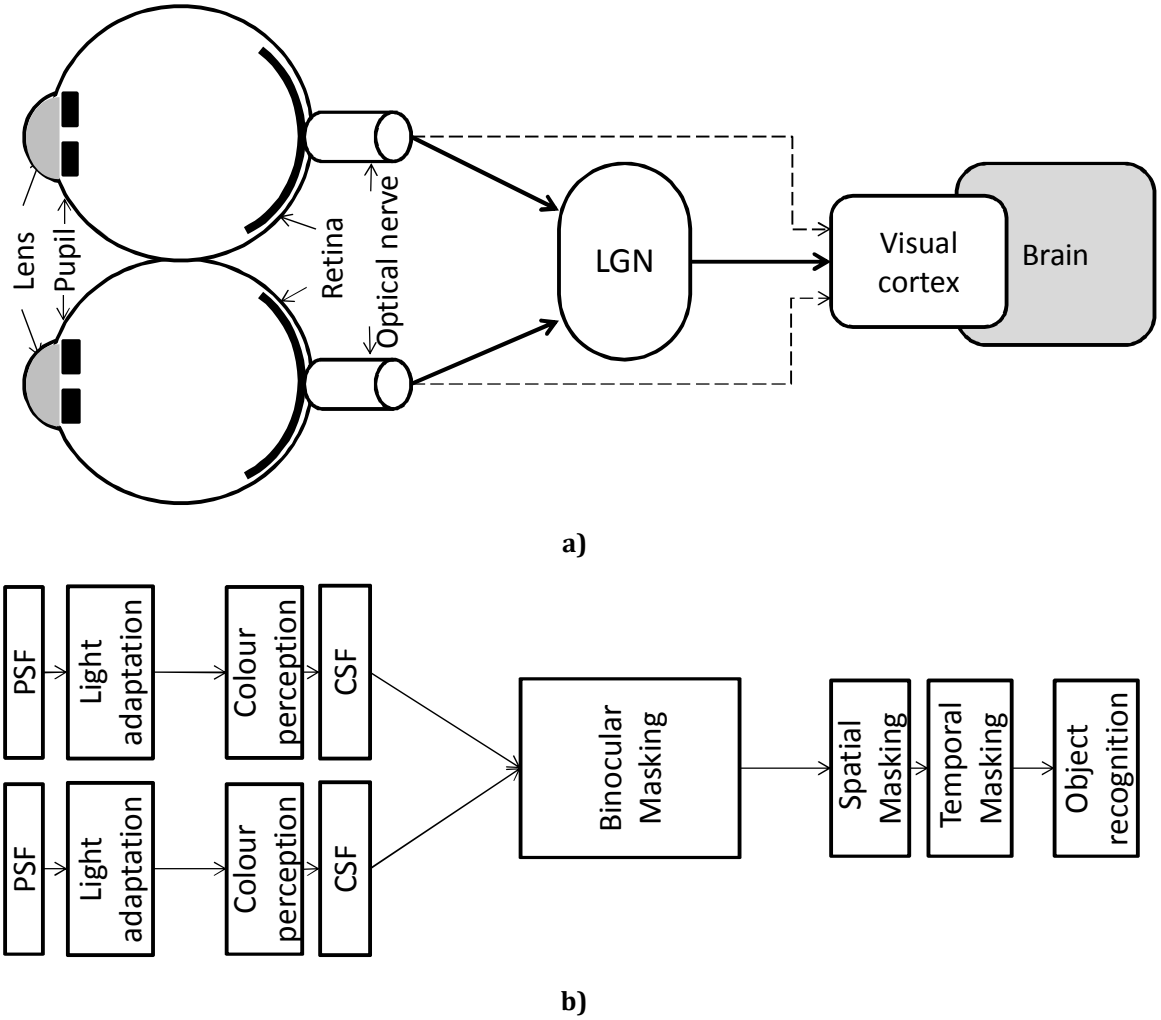


Figure 4. Human visual system: a) structural model, b) functional model

spectral frequencies, with the peak sensitivity at approximately 440nm (blue) for S-cones, 550nm (green) for M-cones and 580nm (yellow-green) for L-cones. The ability of the brain to deduce the colour spectrum using three colour components is known as *trichromaticism*. This ability allows one to construct a full-colour display using a set of monochromatic colour components.

The rods are responsible for the low-light vision and are generally ignored in the HVS modelling [3]. The photoreceptors are not uniformly distributed. Their density on the retina has its maximum at the *fovea* (central point of the projected image), and drops rapidly with increasing distance from it. The retinal area at the spot where the optic nerve leaves the eye does not have

any photoreceptors, resulting in a visual gap commonly known as the *blind spot*. There is sensitivity adaptation mechanism in the retinal cells which works together with the optical system of the eye (the iris controlling the amount of light entering through the pupil). As a result, eyes work over a wide range of luminance values but are sensitive only to relative luminance changes (i.e. contrast), rather than absolute luminance values. This HVS property allows displays with limited luminance output to faithfully represent scenery lit by bright daylight (however, the dynamic luminance range of such displays is limited).

The visual information leaves the eye through the optic nerve, formed by the long axons of the retinal ganglion cells. There are about one million fibres per eye [7]. The fibres of each retina are reorganized in the *lateral geniculate nucleus* (LGN) before being fed to the *visual cortex*. The structure of the LGN suggests two separate visual processing subsystems; one with high temporal and low spatial resolution (localization) and another which has slow response but high resolution in space (identification) [7]. Such separation allows the motion information to be encoded using temporal resolution of as little as 15 frames-per-second.

The functional model of the binocular vision is shown in Figure 4b. The eye can change its refractive power in order to focus on objects at various distances. The process is known as *accommodation*, and the refractive power is measured in dioptres. The imperfections of the optical systems cause a blur to the projected image, which is typically modelled as a low-pass filter characterized by a point spread function (PSF) [8]. The combination of the iris, controlling the amount of light entering the eye, and the sensitivity adaptation of the retina allow the eye to work over a wide range of intensities (between 10^{-6} and 10^{8} cd/m^2). The eye is sensitive to luminance difference (i.e. contrast), rather than absolute luminance values. This visual property is known as *light adaptation* and is modelled by local contrast normalization [9]. The region of the visual fixation point is perceived with the highest spatial resolution; this is called *foveal vision*. The surrounding vision, with rapidly decreasing resolution is known as *peripheral vision*. This effect is usually modelled by re-sampling the image with a non-regular grid (denser in the fixation point and sparser away from it), in a process, known as *foveation* [10]. The trichromatic colour vision is modelled using a perceptual colour space [3] [11]. Due to the way visual information is processed, the HVS has different sensitivity to patterns with different density. The minimum contrast necessary for an observer to detect a change in intensity is called a threshold contrast, and it dependence on pattern density is described by so-called *contrast sensitivity function* (CSF) [3] [8]. The neurons in the visual cortex are sensitive to particular combinations of spatial and temporal frequencies, spatial orientation and directions of motion. This is well-approximated by two-dimensional Gabor functions [3] [8]. The spatially dependent CSF is used for perceptually optimized compression of images [12].

The visual information is collected by the photoreceptors in the retina of each eye separately. The luminance, colour and contrast adaptation occur in each eye separately. After that, both optic nerves arrive at the LGN. The LGN is thought to de-correlate the visual information, greatly reducing the visual information – the number of outgoing visual nerves is only 1% of the number of neurons going to the LGN. The processes of binocular masking and extraction of binocular depth cues happens at that stage. The output of the LGN is a fused representation of the scene which appears as if observed from a point between the eyes. This representation is called *cyclopean image*. The cyclopean image is fed to the V1 visual brain centre. The processes in V1 are modelled as multi-channel decomposition, masking between channels with different spatial frequency and orientation, and finally, temporal sensitivity and masking. The binocular

suppression theory and also anatomical evidence suggest that a small part of the visual information delivered from each eye might be fed directly to V1 without being processed by the LGN.

The ability to perceive visual information through two distinct eyes is known as *binocular vision.* Compared to other types of visual perception, binocular vision appeared late in evolution. In nature the hunting animals tend to have broader field of binocular vision, while the hunted animals have a broader, but monocular field of view [5]. The eyes of a human are separated horizontally by a distance of approximately 6.3cm on average [4]. Such positioning allows each eye to perceive the world from a different perspective, as shown in Figure 5. The observer can control the visual fixation point though the *extraocular muscle system*. A group of six muscles surround each eye, and they work together to move and coordinate the eyes. The eyes are able

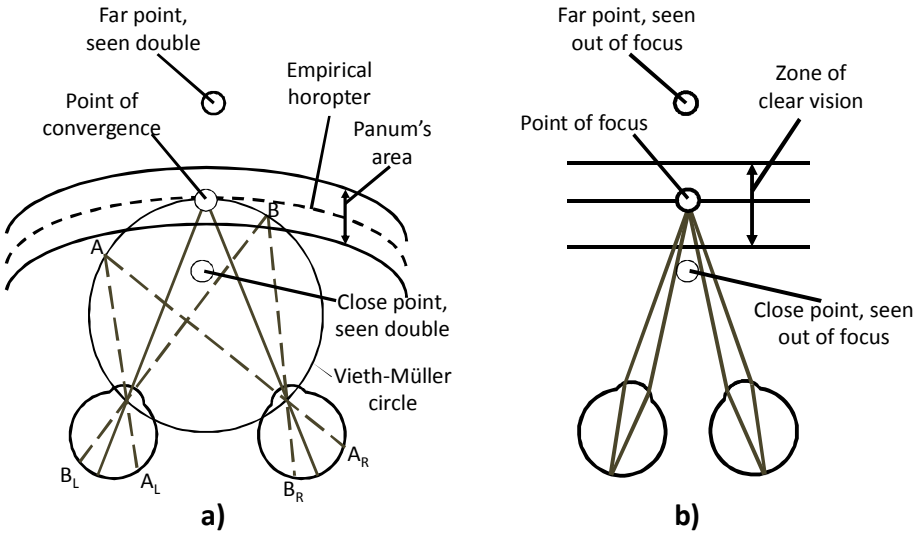


Figure 5. Binocular geometry: a) horopter for a given point of convergence, b) zone of clear vision for a given point of focus

to fix onto an object, and to hold the fixation as the object moves (tracking). They turn inward for near viewing, and are parallel for a distant object. Since eyes perceive the scene from different perspectives, the projections of an object around the point of convergence are not identical. The existence of two different retinal images is called *binocular disparity* [4]. This difference is used by the HVS to deduce information about the relative depth between different points of interest. The ability of the HVS to perceive depth using binocular disparity is known as *stereovision*.

All points that are projected onto identical places in each retina (relative to the fovea) can be fused by the HVS. For a given point of convergence, there are points which are projected with identical offset relative to each fovea; these are shown for points "A" and "B" in Figure 5a. The set of all points which are projected onto matching retinal positions is called *horopter*. The theoretical horopter coincides with the circle which passes through the point of convergence and the centre of each eye's lens as shown in Figure 5a. That circle is also known as *Vieth-Müller circle*. However, the horopter derived trough subjective experiments (also called *empirical horopter*), does not fully coincide with the theoretical one [13]. Around the horopter, there is a region of points, which projections can be fused by the HVS. That region is known as *Panum's area*. Outside of the Panum's area binocular depth is still perceived, but the objects are seen as

doubled. The experience of seeing double objects is known as *diplopia.* When eyes focus on a point the refractive power of each eye changes in order that the projections of that point appear in focus in each retina, as seen in Figure 5b. Close to the point of focus there is a larger area, where objects are perceived in focus. The area is known as *zone of clear vision* and its size depends on the distance to the point of focus, and also on the size of the iris. In order to speed-up the accommodation process the convergence and focus of the eyes are simultaneously driven by the so called *accommodation-convergence reflex*. The distance to the point of convergence influences the focal distance, and vice versa. In a natural 3D scene, such coupling increases the speed of accommodation and helps the convergence process by blurring the objects in front and behind the convergence point.

Further details about the physiology and modelling of the binocular HVS can be found in the following books [6] [3] [8] [14].

2.1.2 Visually important features of a 3D scene

The HVS is a set of separate subsystems, which operate together in a single process. It is known

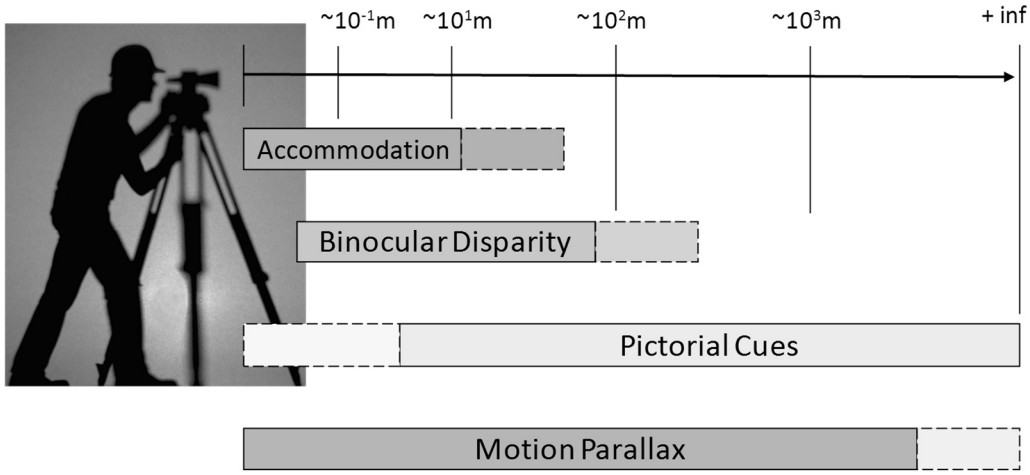


Figure 6. Depth perception as a set of separate visual "layers". Appears in [P08], published by SPIE. © 2009 SPIE, reprinted with permission.

that spatial, colour and motion information is transmitted to the brain using largely independent neural paths [3]. Vision in 3D, in turn, also consists of different "layers" which provide separate information about depth of the observer scene [3] [5]. This is true both for perception (separate visual mechanisms and neural paths) and cognition (various properties of the scene are used for perceiving the depth). The visual features used by the HVS for depth perception are also known as *depth cues*. There are separate families of depth cues, with varying importance from observer to observer [4] [15] [16]. It is possible that masking and facilitation effects exist between depth cues. The presence and strength of one depth cue type might suppress or enhance the perceptibility of another.

The importance of different depth cues varies with the distance as shown in Figure 6. There are the following groups of depth cues:

- *Focal depth* – the HVS can use the refractive power of the eye as a depth cue. In short distances, accommodation is the primary depth cue, since closely positioned objects are hardly visible with two eyes. With increasing observation distance, the importance of this depth cue quickly drops.

- *Binocular depth* – retinal disparity is used as a depth cue providing relative distance. Binocular depth cues are the ones most often associated with "3D cinema". Approximately 5% of all people are "stereoscopically latent" and have difficulties assessing binocular depth cues [3] [5]. Such people rely on depth information coming from other cues.
- *Pictorial cues* – for longer distances, binocular depth cues become less important and the HVS relies on pictorial cues such as shadows, perspective lines and texture scaling for depth assessment. Pictorial depth cues can be perceived by a single eye.
- *Head parallax* (also known as *motion parallax*) – this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. Observers naturally expect to be able to see the scene from different perspectives by changing their head position. The same mechanism is used by insects, and is commonly known as "insect navigation" [17].

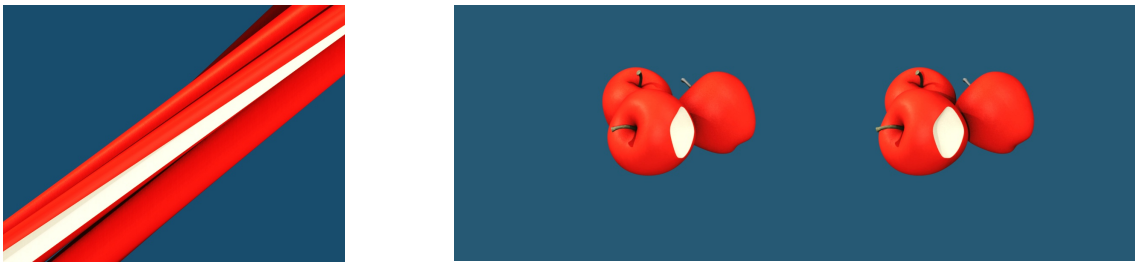
More detailed information about the binocular depth perception can be found in [4] [5] and [18].

2.1.3 3D scene sensing and representation

A 3D scene sensing technique attempts to solve the ill-posed problem of reconstructing a 3D scene from a limited number of remote observations. An overview of 3D sensing techniques is available in [19]. One group of methods works on 3D scene captured by a single camera. Such methods work by analysing monocular depth cues. In this category are *shape-from-shading* [20], *shape-from-texture* [21], *shape-from-defocus* [22] and *shape-from-motion* [23]. Another single camera 3D sensing approach involves fitting a 3D model over known 3D shapes, e.g. face [24] or body [25]. This is equivalent to processes in which the HVS assumes size and 3D shape of known objects. The second group of techniques attempts to reconstruct a scene captured by two or more cameras. The main problems in that approach are finding corresponding features in each observation, and reconstruction of occluded pixels [26]. The third group of methods use active camera sensing, and capture 3D data by projecting *structured patterns* or *coded light*. Another active 3D sensing approach is *time-of-flight* imaging, where the camera emits light signal and measures the time it takes for the signal to reach the scene and bounce back to the camera [27]. Finally, there are holographic 3D scene capture methods which record the interference pattern, created by superimposing a reference beam with a beam scattered by the scene. If the interference pattern is captured by a CCD camera instead of holographic material, the technique is known as *digital holography* [28].

Selection of 3D scene representation format is a compromise between two goals – first, to have of an accurate description of the "important" visual features, and second, to have compact description, which is suitable for storing and transmission. Most formats for representing visual data descend from the human understanding of a natural scene in terms of geometry and texture. However, scene description formats are also greatly influenced by peculiarities of the content creation process – 3D capture for natural scenes and 3D rendering for synthetic ones. While the concrete details in encoding, compression or file structure might differ, there are three major groups of abstract 3D scene representation [29].

The first is so-called *spatio-perspective volume*, where a multiple viewpoints of the same scene are recorded [30]. Such volume is created by capturing or rendering images from different camera perspectives. The camera can move in a 2D plane, and capture full scene parallax, or along a line and capture horizontal parallax only. Due to the similarity between the images seen from neighbouring locations (which is called *perspective coherence* in [30]), the spatio-temporal



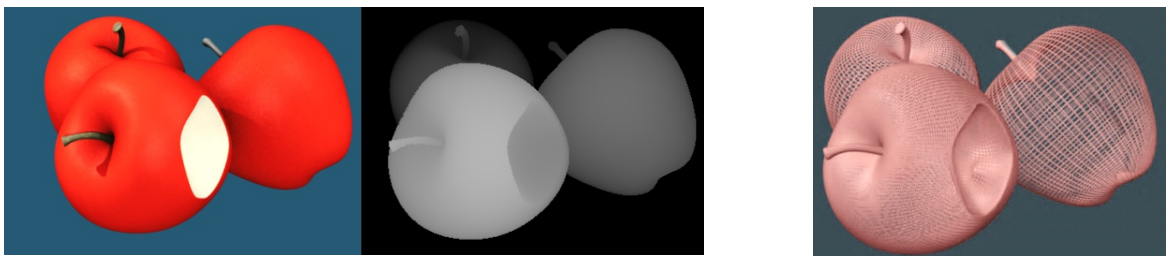
a) b)

Figure 7. Spatio-perspective representations of a 3D scene: a) epipolar image , b) side-by-side stereosopic pair

volume is a description which contains great amount of redundancy. Observations of objects captured by a linearly moving camera appear with linear shifts (a property of a 3D scene known as *epipolar constraint* [31]). As a consequence, a slice of the spatio-temporal volume parallel to the perspective dimension contains many straight lines, as shown in Figure 7a. The lines are known as *epipolar lines* [31] and the slice is known as *epipolar plane image* (EPI) [30].

When sliced across the perspective dimension, the volume contains a number of scene observations from different perspectives (known as *views*). A scene representation which contains a limited number of these observations (typically 2-30) is known as a *multiview image* [29]. A relatively simple way to store a multiview image is to combine all observations in a single bitmap, and to store a multiview image or a stereoscopic pair in a so-called *side-by-side* fashion as shown in Figure 7b. A more sophisticated approach is to encode the differences between the observations similarly to the way temporal similarities are encoded in a video file as done in MPEG-4 MVC [32]. Multiview images are one of the most common 3D scene description formats as they are straightforward to capture or render.

The second group of scene representations is *video-plus-depth*, where each pixel is augmented with information of its distance from the camera. A straightforward way to represent video-plus-depth is to encode the depth map as a grey scale picture, and place the 2D image and its depth map side-by-side. The intensity of each pixel from the depth map represents the depth of the corresponding pixel from the 2D image. Such a format is sometimes referred to as *2D+Z*. An example of 2D+Z representation of a scene is shown in Figure 8a. Video-plus-depth format is suitable for multiview displays as it can be used regardless of the number of views a particular screen provides [33] [34]. Furthermore, video-plus-depth can be efficiently compressed. Recently, MPEG specified a container format for video-plus-depth data, known as MPEG-4 Part-3 [32]. On the downside, video-plus-depth rendering requires disocclusion filling, which can be a source of artefacts. This is being addressed by using layered depth images (LDI) [29] or by



a) b)

Figure 8. Representations of a 3D scene: a) 2D+Z, b) Mesh

multi-video-plus-depth encoding [35]. Visualization of 2D+Z video on a multiview display requires additional computation. Based on the depth map provided with the scene, multiple observations should be rendered and the pixels from these observations should be interleaved in the way required for the display. The dense depth map is not captured directly. It can be derived from multiview images (using depth estimation algorithms) or from point cloud data captured by range sensors. In the case of a synthetic 3D scene, obtaining a dense depth map is a straightforward process, as solving the occlusions during rendering requires calculation of the distance between camera and each pixel of the image [36].

The third group of representations store the scene geometry in a vectorized form. One example is the dynamic 3D mesh [32]. Such representation is suitable for synthetic content, since synthetic 3D scenes are originally described as form. Examples of mesh representation is shown in Figure 8b. Mode details on 3D scene representation formats can be found in [29] and [32].

2.2 3D displays

Three-dimensional displays are ones which can show a visually indistinguishable copy of a real 3D scene. The ideal 3D display would recreate all depth cues of a scene, regardless of their importance or applicability in the usage scenario. Due to design constraints, a real 3D display can show only a sub-set of the required depth cues, and the cues that are present are often

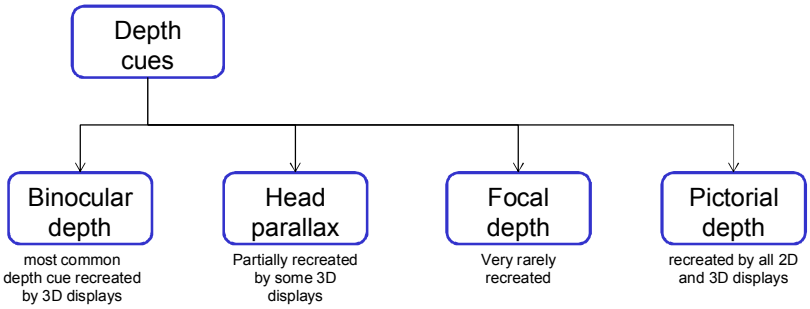


Figure 9. Depth cues, recreated by 3D displays.

partially recreated. The name "3D display" is often used as a marketing term for a display which can generate any additional depth cues in comparison to the older generation of "2D displays".

In Figure 9 one can see a list of depth cues created by various modern displays. Most often, a display earns its "3D" label by being able to provide a separate image for each eye of the observer. Due to scene parallax, objects appear on different coordinates in each image. The offset between the observations is known as *image disparity*. In a stereoscopic pair, image disparity leads to binocular disparity and creates the illusion of depth. In this thesis, the illusory distance to the object created by the stereoscopic effect is called *apparent depth*. Positive disparity creates apparent depth behind the screen plane, and negative disparity creates apparent depth in front of the screen.

Most contemporary 3D displays do not recreate head parallax. Some models can present limited head parallax by casting different images towards a set of observation angles, usually limited to a horizontal head parallax only. Note, that by using head-tracking it is possible to present a scene from different perspectives on a monoscopic display, thus generating head parallax without binocular depth cues [37]. Focal depth cues are very rarely recreated by 3D displays. One exception is the stereo display prototype with multiple focal distances described in [38]. Finally, pictorial depth cues can be recreated by most 2D and 3D displays (volumetric LED cube

displays [39] being an exception). More information about various types of 3D displays can be found in [40] [41] [42] [43].

2.2.1 Classification

There are a number of taxonomies of 3D displays. A general one divides them into three basic types: holographic, volumetric and multiple-image screens [41] [44]. Holographic displays use holographic methods to reconstruct the light field of a scene, volumetric displays attempt to approximate a 3D scene by light elements (voxels) positioned in 3D space and multiple image screens cast a number of different images, each one seen from a different angle. There are two types of multiple-image screens. The first type works by tracking the observer's eyes, and utilizes steerable optics to beam different images towards each eye. The second type uses fixed optics, and beams a number of views in different directions; the directions are selected in such a way, that the eyes of an observer standing in front of the screen perceive different images. In

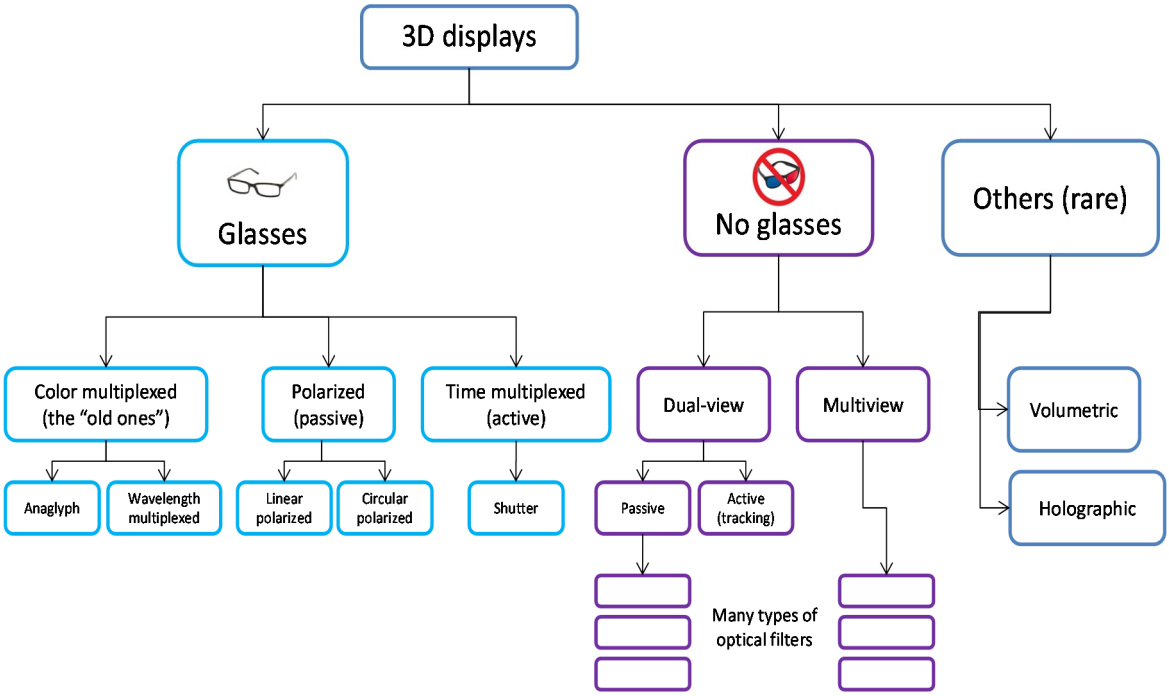


Figure 10. Classification of 3D displays

[40], these two types are said to create *eye-gaze-related images* and *fixed-plane images* correspondingly. The taxonomy in [41] is different – displays with steerable optics are named "*head position tracking displays*", while the ones with fixed optics are designated simply as "*multiview displays*". This dissertation follows the terminology in [41], and uses *multiview display* to designate autostereoscopic display which generates multiple images by means of fixed optics.

The classification used in this thesis is shown in Figure 10. Instead of following the methods for image creation, it classifies the 3D displays taking the observer's perspective. For the user point of view, the main differentiation factor is whether the display requires glasses or not. Thus, the taxonomy in this thesis has "glasses-based" or" glasses-free" as major display types. The predominant share of 3D displays in the market is binocular stereoscopic TV sets, which use a thin film transistor liquid crystal display (TFT-LCD) for image formation, and require the observers to wear glasses. Colour multiplexed anaglyph glasses are rare, though some 3D

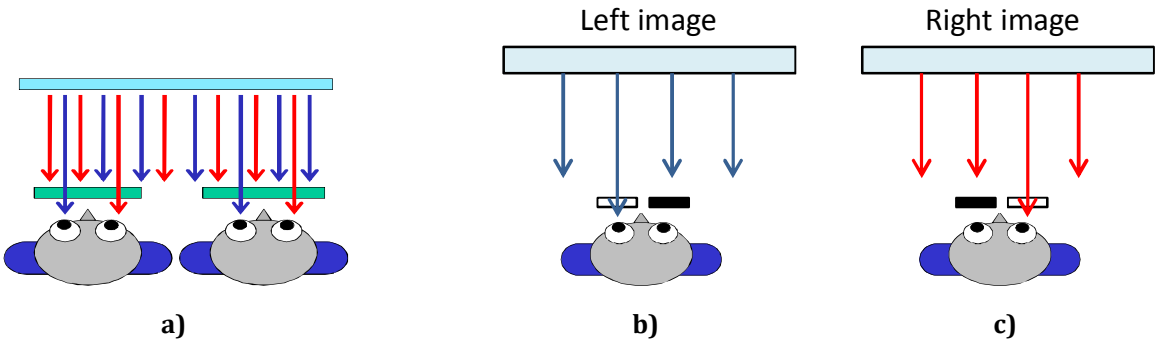


Figure 11. Glasses-based 3D displays: a) general principle of operation, b)-c) operation principle of temporarly interleaved glasses, b) left view visible, c) right view visible.

cinemas still use wavelength multiplexed glasses [45]. The 3D TV sets are sold either with *polarized glasses* (marketed as "passive") or *temporally-multiplexed* ones (marketed as "active"). The displays without glasses are separated into two groups – binocular autostereoscopic ones, mostly used in mobile devices, and multiview displays, used for outdoor advertising or (rarely) in computer setups. As an exception, Toshiba announced a 3D TV model which uses a combination of a multiview display and observer-tracking [46]. All other types of 3D displays, for example volumetric or holographic, are rare. Notably, most 3D displays marketed as "digitally holographic" do not use holographic visualization techniques, but are, in fact, multiview displays.

2.2.2 Glasses-enabled stereoscopic displays

Glasses-enabled 3D displays use one display surface to beam two views (one for each eye). Each view can be seen from a range of observation directions. Glasses worn by each observer separate the light beams, so each eye receives only the intended view, as shown in Figure 11. Temporally-interleaved 3D displays beam both views, alternating them over time. The observer wears active glasses, which work synchronously with the display and block the light to one or the other eye at the correct time. When the display is beaming the left image, the light towards the right eye is blocked (see Figure 11b), and when the right image is beamed, the light to the right eye is blocked (see Figure 11c). At any instant of time only one of the observer's eyes perceives the image, however due to the high speed of the process (120-240 frames per second), the user is unaware of the temporal interleaving.

Another approach is to beam both images using differently polarized light, and use polarization filters in front of each eye. In this case, each eye receives differently polarized light, but since the HVS is not sensitive to light polarization, the observer is unaware of the polarization-based separation. More frequently, circular polarization is used (clockwise for one eye and counter-clockwise for the other) which allows the beam separation to work for a wide range of head orientations (e.g. head tilt). Passive polarizing glasses are used with both light-emitting TV displays (see Figure 12a) and light-reflecting projector-based displays (see Figure 12b). Projector-based setups use two projectors equipped with polarizing filters and require a special reflecting surface in order to preserve the polarization of the reflected light. Since two projectors are used, each eye receives image with the same resolution. The light emitting stereoscopic displays with passive glasses (hereafter abbreviated as *SDPG*) use spatial interleaving. In such displays, the available TFT elements are divided into two groups with different polarization, as shown in Figure 12c. Each group is visible to one of the eyes only. As in

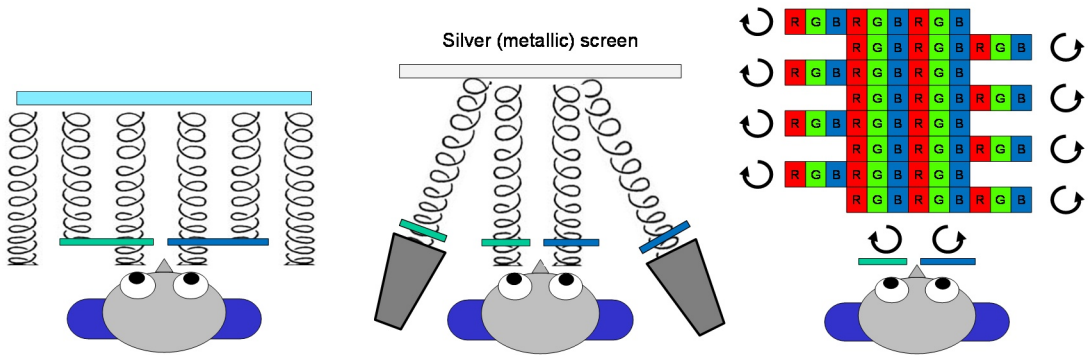


Figure 12. 3D displays with passive glasses: a) light-emitting display, b) light-reflecting display and c) row-interleaving, used with light-emitting displays

binocular displays horizontal resolution is more important that the vertical. The groups are usually row-interleaved – the rows with odd numbers are visible by one eye and the rows with even numbers by the other. Each eye sees the other half of the rows dark – for example the left eye may see the image in the odd rows, and black stripes in the place of the even rows.

2.2.3 Dual-view autostereoscopic displays

Dual-view autostereoscopic displays beam two images with each one seen from a different perspective. Usually, each image can be seen from a number of observation angles as shown in Figure 13a. This allows a number of observers to use such display, provided that each observer is correctly positioned. A practical example of positions where one of the views is visible is shown in Figure 13b. The figure is a photograph of a dual-view autostereoscopic display, beaming two images – one ”white” image where all pixels are at full brightness, and another “black”, where all pixels are off. On the figure one can see where the “white” image is visible.

There are a number of designs which allow one display to beam two different images. The most common approach is to put an additional layer in front of the TFT-LCD [40] [41] [47]. TFT displays recreate the full-colour range by emitting light though red, green and blue coloured components (*sub-pixels*), usually arranged in repetitive vertical stripes as shown in Figure 14. The layer alters the visibility of each pixel, and makes only half of the sub-pixels visible from a given direction. The layer is called “*optical layer*” [48], “*lens plate*” [40] and “*optical filter*” [49]. The design, where only part of the sub-pixels is visible from a given direction is also known as

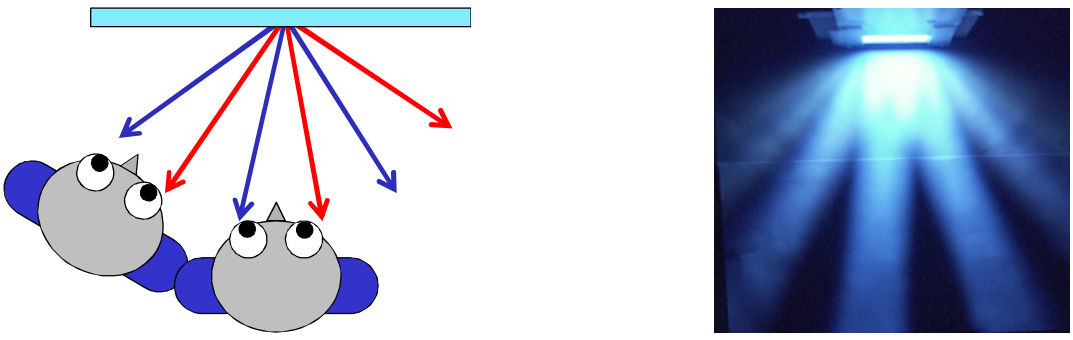


Figure 13. Dual view autostereoscopic displays: a) general principle, b) positions, where one of the views is visible. (b) appears in [P05], reprinted with permission.

spatially-multiplexed autostereoscopic display [40].

There are two common types of optical filters; *lenticular sheet* and *parallax barrier*. Lenticular sheets are composed by small lenses which refract the light to different directions as shown in Figure 14a [48]. A parallax barrier is essentially a mask with openings and closings that blocks the light in certain directions shown in Figure 14b [47]. In both cases the intensity of the light rays passing through the filter changes as a function of the angle, as if the light is directionally projected. Also, as only half of the available sub-pixels belong to each of the views, the resolution of each view is lower than the full 2D resolution of the display.

One way to provide each view with the full resolution of the display is to use temporal interleaving. One example is the 3D display with the patterned retardation film produced by 3M. It distributes the light into two perspective views in a sequential manner, as shown in Figure 14c. The display uses a standard TFT panel and two separate backlighting sources. The two backlights are turned on and off in counter-phase so that each backlight illuminates one view. The switching is synchronized with the LCD, which displays different-perspective images at each backlit switch-on time. The role of the 3-D film is to direct the light coming from the

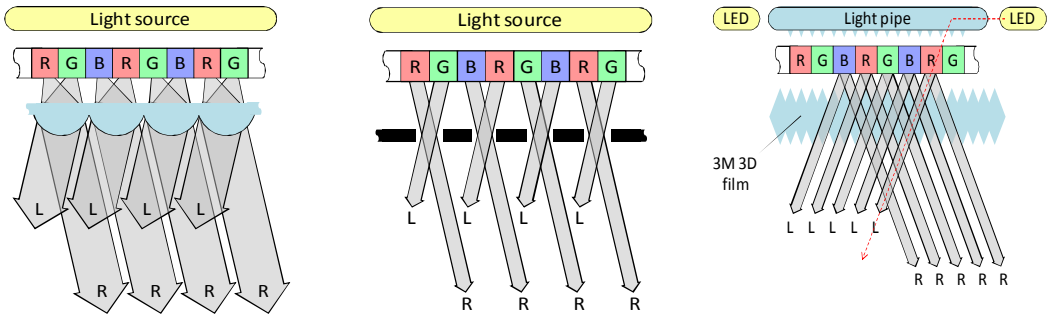


Figure 14. Optical filters for autosereoscopic displays: a) Lenticular sheet, b) parallax barrier and c) temporaly interleaved patterned retarder. Appears in [P02], © 2010 IEEE, reprinted with permission.

activated backlight to the corresponding eye. More information on autostereoscopic displays can be found in [43] [50] and also in [P02], included in this thesis.

2.2.4 Multiview displays

Most multiview 3D displays work in a similar fashion to the spatially-multiplexed dual-view ones. However, in contrast to having their sub-pixels separated into two views, multiview displays have more views, typically 8 to 24. The current generation of multiview displays uses the same basic principles for light distribution - lenticular sheets [51] or slanted parallax barrier [49]. The lenticular sheet works by refracting the light as shown in Figure 15a, and the parallax barrier which works by blocking the light in certain directions as shown in Figure 15b. In both cases the intensity of the light rays passing through the filter changes as a function of the angle [48]. Since sub-pixels appear displaced in respect to the optical filter their light is redirected towards different positions. As a result, differently coloured components of one pixel belong to different views. Respectively, the image formed by one view will be combination of colour components (sub-pixels) of various pixels across the TFT screen. When red, green and blue sub-

pixels are visible from the same direction and appear close to each other, the triplet is perceived as one full-colour pixel. Such pixel is a building block of the view seen from that direction.

As a result of applying the optical filter, for every sub-pixel there is a certain angle from which it is perceived with maximal brightness – that angle is called the *optimal observation angle* for the sub-pixel. The vector, which starts from the sub-pixel and follows the optimal observation angle, is the *optimal observation vector* for the sub-pixel. The optimal observation vectors for all sub-pixels of the same view are designed to intersect in a tight spot in front of the multiview display. From this spot, the view will be perceived with its maximal brightness. In this thesis, that spot is referred to as being the *optimal observation spot* of the view. Outside of the optimal observation spot there is a range of observation angles, from which a given view is still visible, even though with diminished brightness. In this text such range is called the *visibility zone* of a view. For most multiview displays visibility zones of the views are ordered in horizontal direction. A notable exception is the SynthaGram display produced by StereoGraphics [52] which has 9 views with visibility zones ordered in 3-by-3 grid. As the amount of the pixels provided by the underlying TFT is limited, there is a trade-off between the number of views casted by a 3D display and the resolution of each view. As stereoscopic depth cues are generally perceived in the horizontal

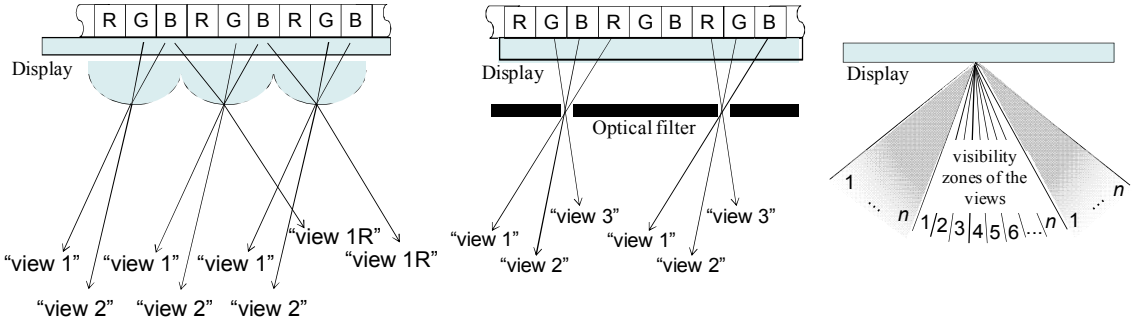


Figure 15. Multiview displays: a) lenticular sheet, b) parallax barrier, and c) visibility zones of the views. (a) appears in [P10], (b) (c) appear in [P03], reprinted with permission.

direction, most multiview display designs do not allocate pixels for extra views in vertical direction [41] [49] [50] [51].

When horizontally ordered, the visibility zones appear in a fan-shaped configuration as depicted in Figure 15c. The repetitive structure of the optical filter creates several observation zones for any view, which follow the fan-shaped configuration as well. After the visibility zone of the last view, the first view becomes visible again. This creates one central set of visibility zones directly in front of the screen, and a number of identical sets to the side as shown in Figure 15c. An example for two observation angles from which the same set of sub-pixels is visible is shown in Figure 15a. The observation angles marked as "1" and "1R" are optimal observation angles of the same view.

2.2.5 Autostereoscopic displays modelled as signal processing channel

One original idea proposed in this dissertation is a general model of spatially-multiplexed 3D displays, which considers it as a signal processing channel. Such a model allows one to relate the optical properties of a 3D display to its visual quality. It also helps to describe artefacts in 3D displays in terms of signal processing, which allows one to address artefact mitigation with

signal processing methods. The model, proposed in this dissertation, extends the ideas for topological description of the interdigitation map proposed in [52], and for describing individual horizontal lines in terms of ray-space samples proposed in [53].

A model, describing a spatially-multiplexed dual-view 3D display is presented in Figure 16. The model is applicable to SDPG displays and dual-view autostereoscopic displays. The model starts with two continuous signals that represent input to each of the two channels. In Stage I, each signal is sampled at the full resolution of the display. In Stage II, each channel is decimated by a factor of two, which models the fact that only half of the sub-pixels are visible to each eye. For the case of a SDPG display, the decimation is in vertical direction which models the fact that only half of the rows are visible to each eye. For a dual-view autostereoscopic display, the decimation would be in horizontal direction as these displays are column interleaved. In Stage III an upsampling block represents the fact that half of the rows are seen as dark stripes. The sampling rate is increased two times and zeroes are inserted between the samples. In Stage IV the samples from both channels are interleaved, which is equivalent to the sum of the two upsampled channels where one channel is delayed by one sample. The final stage models the

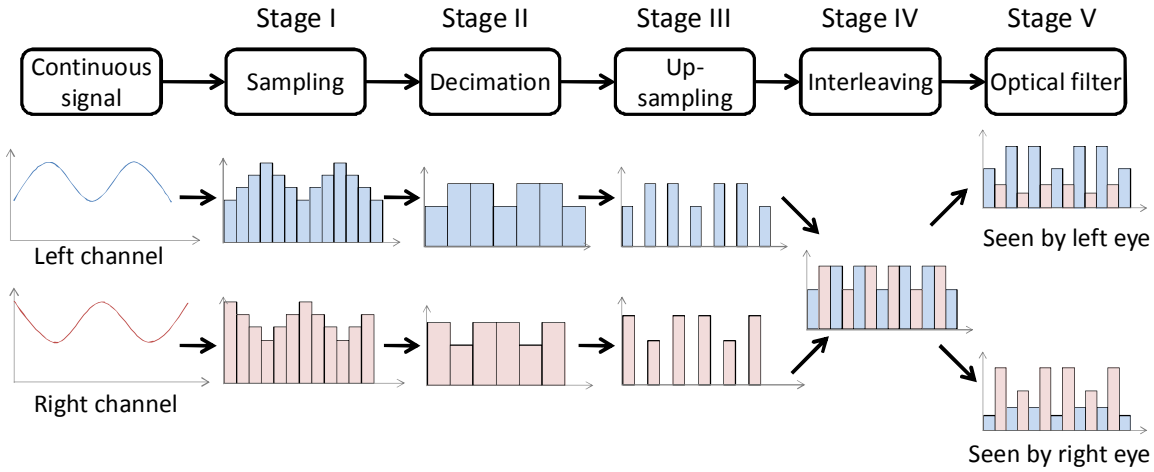


Figure 16. Model of a spatially-multiplexed 3D display with passive glasses as an image processing channel.

effect of the optical filter. Half of the samples (the ones supposed to be visible) are left unchanged, and the other half (the ones that should be suppressed) are multiplied by a coefficient between 0 and 1, which represents the effect of the optical filter. It is possible that part of the light intended for one eye is visible to the other. This effect can be due to non-ideal light separation in the optical filter (for autostereoscopic displays) or less-than-ideal polarization filters (in SDPGs).

Similar stages are present in the model of a multiview display, as shown in Figure 17. The first part of the model is the process where the sub-pixels of the views are rearranged into one compound bitmap. Such process is also known as *interdigitation*. The input comes from n images, and each image is with the full ("2D") resolution of the display. From each input image only sub-pixels which belong to one of the views are used. This is modelled by a 2D down-sampling operation. Since the views are spatially-multiplexed, each image gets sampled with different horizontal and vertical offset. The image offset is modelled as a signal delay. As the grid of each view has its own offset, each offset is represented by a different delay block with a signal delay of z^{-n_v}, where v is the view number. On the display the sub-sampled image is

represented in its original size. The visible sub-pixels appear either surrounded by black stripes by the parallax barrier, or enlarged by the lenticular sheet. This effect is modelled as an up-sampling stage where the introduced samples are either set to zero, or are repetition of the same sample value. The optical layer of a multiview display acts as directionally selective filter and applies angular luminance function to each sub-pixel of the display. The angle at which the angular luminance has its peak value determines the optimal observation direction of the sub-pixel. This angle is different for each sub-sampled image. The compound bitmap map can be represented as a set of non-overlapping lattices where each lattice contains sub-pixels from a single view only [52]. On an image with the full resolution of the LCD matrix, each of these lattices acts as a rectangular sub-sampling pattern with a different offset. An example is shown in the bottom of Figure 17 where the intersecting dotted lines mark the position of the LCD sub-pixels: one lattice is marked with circles and another is marked with crosses. The last part of the model represents the effect of the optical layer. The impact of the layer on the brightness of the underlying sub-pixels is modelled as *visibility* - the ratio between the relative brightness of a view and the maximum brightness of the display as seen from the same angle. The visibility of each view is a function of the observation angle. The function gives the visibility of a given view from observation angle θ. The model uses the assumption that the function is the same for all views, with the peak visibility of each view occurring at different observation angle. In Figure 17, k_v is used to denote the angular offset in the visibility function for view v.

More information on modelling of autostereoscopic displays as signal processing channels can be found in [52] [54] [55], and also in [P02] included in this compound thesis.

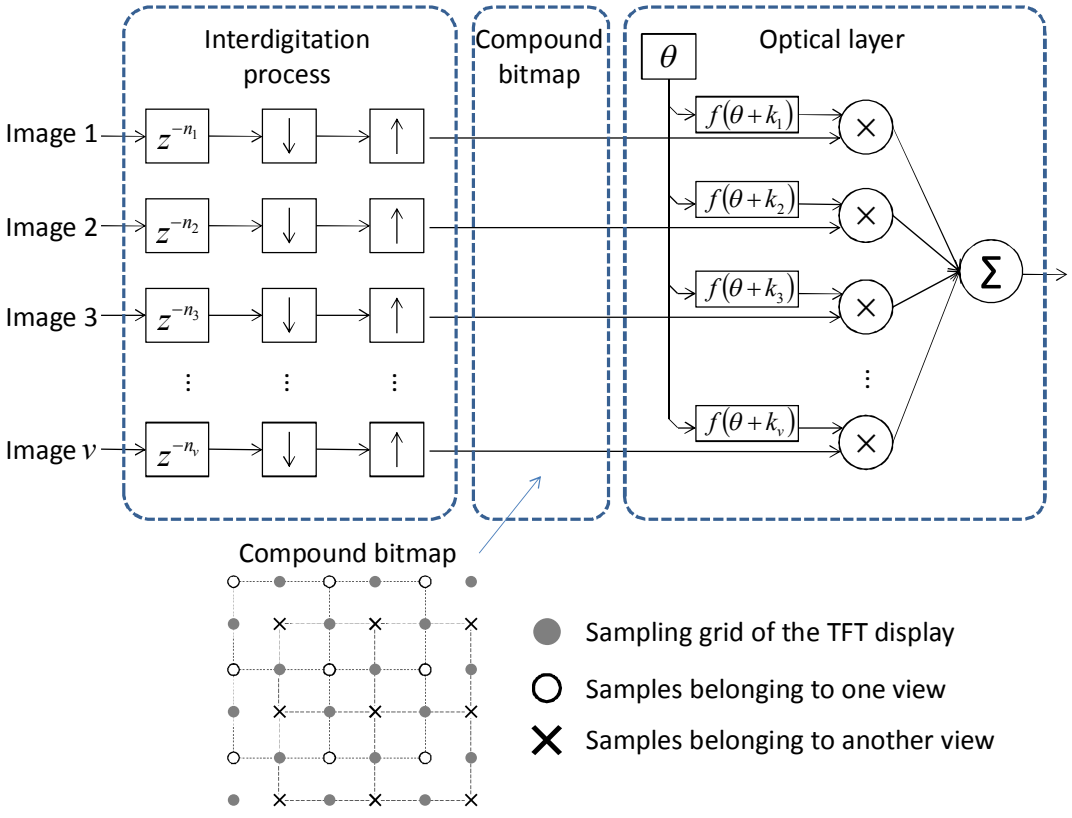


Figure 17. Model of a multiview display as an image processing channel.

3 Visual quality of stereoscopic displays

The visual quality of a 3D display is defined by its capability to visualize 3D scene without visible distortions. Most often, the display is used in a so-called *no-reference* setup – i.e. the observer cannot see the original scene to compare it with its (possibly) distorted replica shown on the display. In the general case of no reference, the visual quality is determined by the presence of recognizable distortions (e.g. artefacts) and the subjective level of annoyance they cause. In this work the distortions are studied according to their origin. They are separated into three large groups, as shown in Figure 18.

The visibility of *viewpoint-related* distortions depends on the position of the observer with respect to the 3D display. Examples of such distortions are ghosting, pseudoscopy and accommodation-convergence rivalry. Viewpoint-related artefacts are common for autostereoscopic displays since the image generated by such displays is a function of the observation angle. However, SDPG displays are also affected by viewpoint-related artefacts since performance of the polarization filter depends on the observation angle. Accommodation-

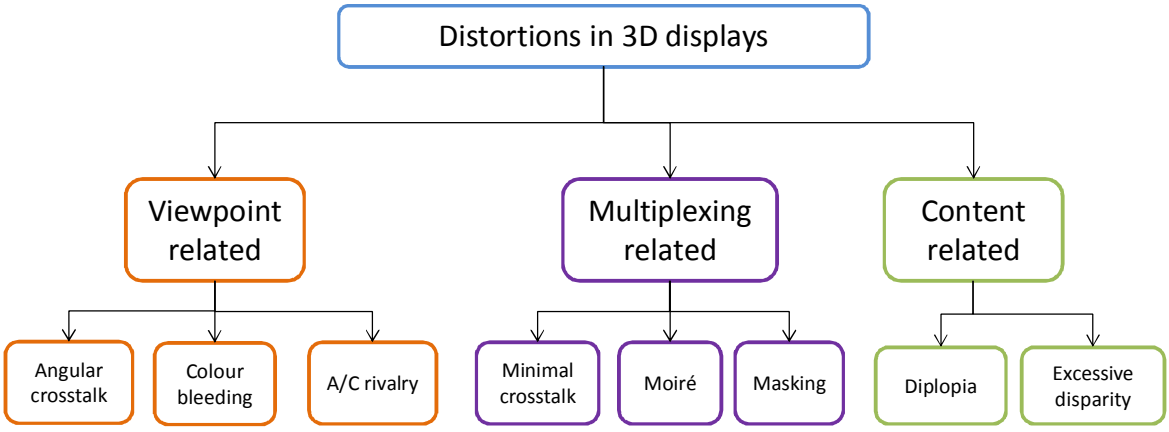


Figure 18. Classification of the most common visual distortions in 3D displays.

related artefacts affect all 3D displays which cannot create images with varying focus.

Multiplexing-related distortions are generated during the process of combining multiple images for presenting onto one display. Sub-optimal channel separation results in some minimal crosstalk regardless of the observation position. Minimal crosstalk affects both temporally and spatially multiplexed 3D displays. Incorrectly prepared images for spatially-multiplexed displays could exhibit Moiré artefacts due to aliasing. Visible gaps between the sub-pixels or non-rectangular pixel shape manifests itself as masking artefacts (also known as fixed-pattern noise).

Finally, some artefacts are caused in the process of content preparation. It is possible that parts of the stereoscopic image are not fuseable. There are two reasons for this; one is that the disparity is too large and the other is that regions of the scene are close to the frame and are present in one channel only. If the observer tries to focus on such an area, he or she experiences *diplopia*. If that happens for objects with an apparent position in front of the screen it is perceived as *frame violation* artefact. Frame violation is more annoying than diplopic objects behind the screen.

3.1 Visibility of image distortions

Webster's Encyclopaedic Dictionary describes *artefact* as "[...] any feature that is not naturally present but is a product of an extrinsic agent or method" [56]. Non-natural processes, as is the case of capturing, coding and recreating a 3D scene, are a source of artefacts. Artefacts are cognitive phenomena. A visual artefact is a deviation from the natural expectation for a scene which is large enough to be perceived. However, it is the cognitive process which recognized the distortion as being an artefact. In absence of reference, the observer cannot estimate the perceptual differences between the original and altered image. Instead, the observer grades the

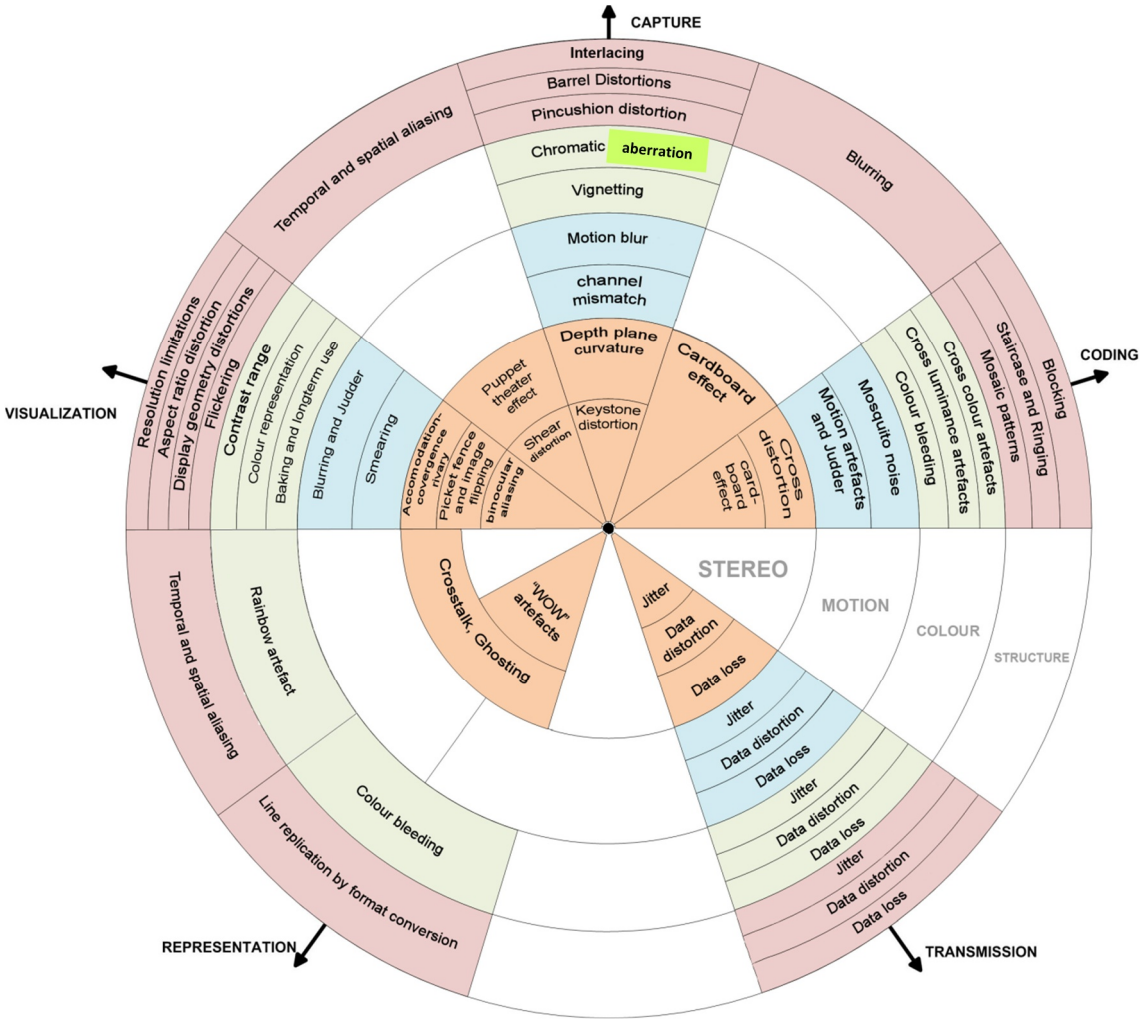


Figure 19. Polar diagram of artefacts in 3D content, ordered by processing stage (angle) and HVS subsystem (distance). Appears in [P08], published by SPIE. © 2009 SPIE, reprinted with permission.

cognitive differences between visible objects and expectation of their "natural" appearance recorded in the visual memory.

Experiments with so-called "random dot stereograms" show that binocular and monocular depth cues are independently perceived [18]. Furthermore, the first binocular cells (cells that react to a stimulus presented to either of the eyes) appear at a late stage of the visual pathways, [3]. This observation leads to the assumption that "2D" (monoscopic) and "3D" (stereoscopic) artefacts would be independently perceived [57]. The planar "2D" artefacts, such as noise, ringing, etc., are thoroughly studied in the literature [58] [59]. The classification in Figure 19

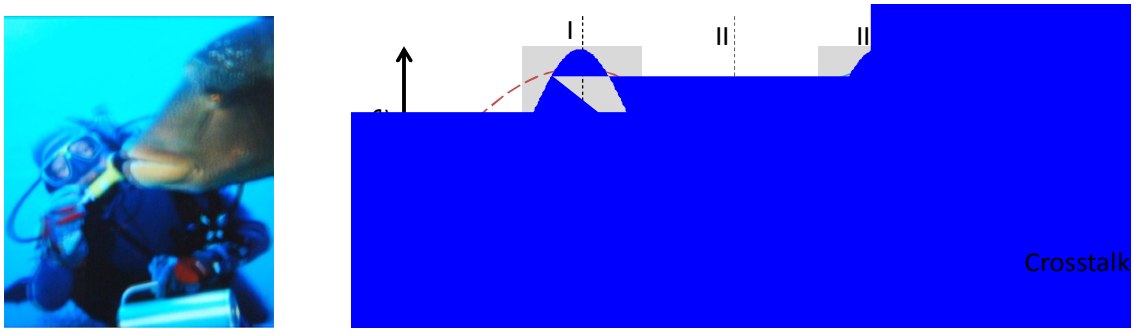
focuses on artefacts which affect stereoscopic perception. However, due to the "layered" structure of the HVS, binocular artefacts might be inherited from other visual subsystems – for example, *blockiness* is a monoscopic artefact, which still can destroy a binocular depth cue.

In this thesis, the taxonomy of stereoscopic artefacts is based on a top-down approach; the processing stages in 3D content delivery are identified, and for each stage one tries to predict these artefacts are interpreted by the HVS subsystems. The taxonomy is presented as a polar diagram in Figure 19. The vector angle represents the processing stages – capture, representation, coding, transmission and visualization. The vector length represents visual subsystems arranged as concentric discs – structure, colour, motion and stereopsis. These layers roughly represent the visual pathways as they appeared during the successive stages of evolution. The spatial, colour-less vision is labelled as *structure*. It is assumed that during the evolution human vision adapted for assessing the "structure" (contours and texture) of images [9], and some artefacts manifest themselves as affecting image structure. *Colour* and *motion* rows represent the colour and motion vision accordingly. All artefacts in the diagram affect the binocular depth perception, however, the row designated with *binocular* contains artefacts which have meaning only when perceived as a stereo-pair. In other words, these are artefacts that cannot be perceived with a single eye (e.g. vertical disparity).

More information on 3D artefacts and their taxonomy can be found in [15] [60] [57] and also in [P01] [P02] [P08] and P10 included in this compound thesis.

3.1.1 Viewpoint-related distortions

If two views are simultaneously visible by the same eye the effect is regarded as crosstalk between the views. If an object of the scene is meant to have apparent depth, its representations in each channel have horizontal disparity. The combination of crosstalk and disparity creates a horizontally-shifted, semi-visible replica of the object. The combination of double contours and transparency is interpreted by the HVS as *ghost images*, or ghosting artefacts [60]. An example of ghost images is shown in Figure 20a. If the amount of crosstalk is different for each colour channel, the shifted replicas have different colours. This effect is referred to as *colour-bleeding*. An example of colour bleeding can be seen in Appendix I, Figure 47. In autostereoscopic displays the visibility of a view is a function of the observation angle, as shown in Figure 20b. The position where one view has maximum visibility, and the other is maximally suppressed is known as the *sweet spot* of that view (marked with "I" and "III" in the figure). The observation

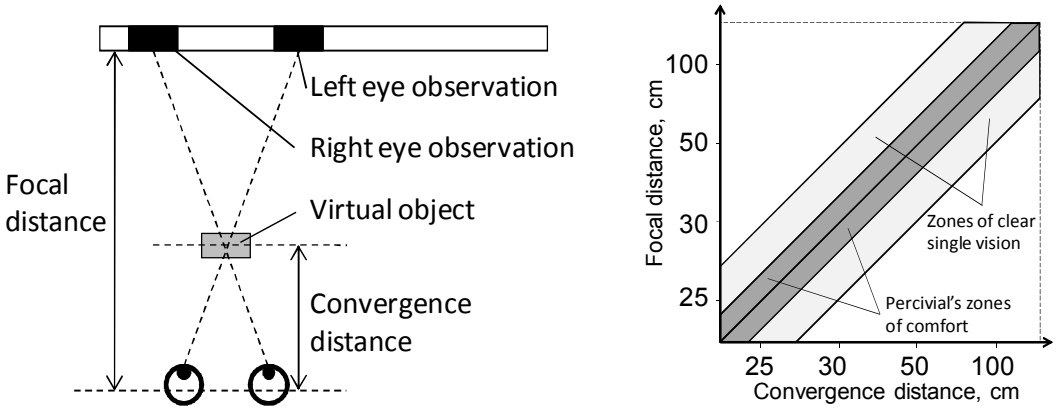


a) **b)**

Figure 20. Ghosting artefacts: a) photograph of a 3D display exhibiting ghosting artefacts (detail), b) crosstalk as function of the observation angle in autosterescopic displays. Adapted from [P05].

zones of the two views are separated by a zone where neither of the views is predominantly visible. That zone is also known as the *stereo-edge* and is labelled as "II" in the figure. For dual-view autostereoscopic display the visibility of the ghosting artefacts is proportional to the crosstalk and has its minimum in the sweet spots and its maximum in the stereo-edge. Subjective visual quality experiments described by Kooi [61] and Pastoor [15] suggest that inter-channel crosstalk of 20% is the maximum acceptable in stereoscopic images.

Another viewpoint-related distortion is the so-called *accommodation-convergence (A/C) rivalry*. On a stereoscopic display the distance to the convergence point can be different from the focal distance, as shown in Figure 21a. This difference is known as *accommodation-convergence mismatch.* The accommodation-convergence reflex drives the eyes to focus at a wrong distance, which causes the objects with pronounced apparent depth to be perceived out-of-focus. Large discrepancy between focal and convergence distance prevents eyes from converging, causing *diplopia*. Stereoscopic fusion is possible only for some combinations between focal distance and convergence distance. The set of focal and convergence distances which allow fusion define so-called *zones of clear single vision*, as seen in Figure 21b [62]. Inside the zones of clear single

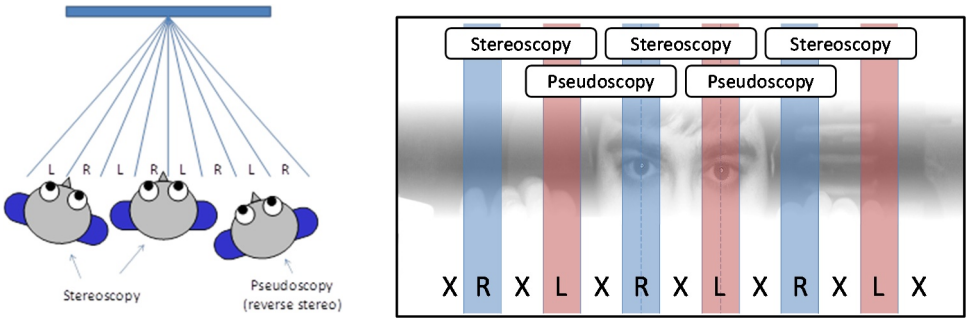


a) **b)**

Figure 21. Acommodation-convergence rivalry: a) focal and convergence distance mismatch, and b) zones of clear single vision and Percival's zones of comfort (adapted from [54]). Appears in [P05], published by SPIE. © 2009 SPIE, reprinted with permission.

vision resides a narrower area, known as *Percival's zone of comfort*, where the difference between the apparent and actually focal distance is less than 0.5 dioptres. Within the Percival's zone of comfort A/C rivalry is negligible [4] [62].

Pseudoscopy (*reverse stereo*) is the situation in which the eyes see the opposite views, i.e. the left eye sees the right view, and vice versa. For example, the two leftmost observers in Figure 22a see proper stereo image while the observer in the right experiences pseudoscopy. In a pseudoscopic image the binocular depth cues contradict the pictorial ones, which results in perceptually disturbing image [60]. Another factor which narrows the size of the sweet-spots is the stereo-edge. Between the stereoscopic and pseudoscopic areas there are there are zones with high crosstalk where 3D effect is not visible; these are marked with "X" in Figure 22b.



a) **b)**

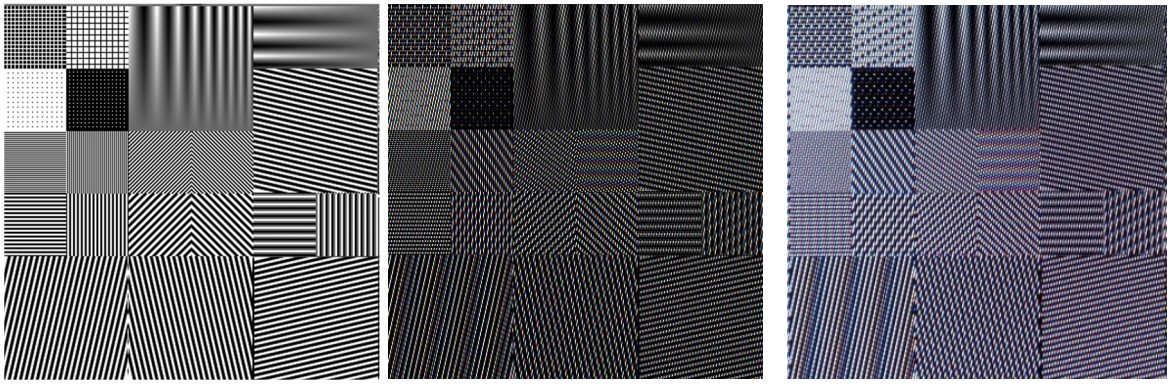
Figure 22. Prseudoscopy: a) stereoscopic and pseudoscopic observation zones viewed from above; b) observation zones yielding clear stereoscopic images.

In addition, some artefacts are most obvious for moving observer: for example the Moiré-like pattern seen on an autostereoscopic display exhibiting *picket fence effect*, or *banding* [60] [48]. Unnatural representation of image parallax causes *shear distortion* in dual-view displays, and *image flipping* in multiview displays [60]. More information about viewpoint-related distortions is available in [15] [60] [61] [62] and also in [P01] [P02] [P05] [P08] included in this compound thesis.

3.1.2 Distortions, related to spatial view multiplexing

In spatially-multiplexed displays the optical filter introduces selective masking over the sub-pixels of the display, thus separating them into different visual channels. This masking can be modelled as a sub-sampling on a non-orthogonal grid. Without pre-filtering this process creates aliasing artefacts which are perceived as Moiré artefacts.

In multiview displays Moiré is visible in all types of scenes, but is especially pronounced in 2D content, as in 3D images aliasing is somewhat masked by more severe artefacts such as ghosting [63]. Visual examples of Moiré artefacts are shown in Figure 23. Figure 23a shows a test image which contains various image details that are susceptible to aliasing. By knowing which sub-pixels are going to be masked by the optical layer one can simulate the output of a multiview 3D display. Such simulation is presented in Figure 23b. Finally, Figure 23c is an actual photograph of a 3D display showing the test image from Figure 23a. The design of the display includes a



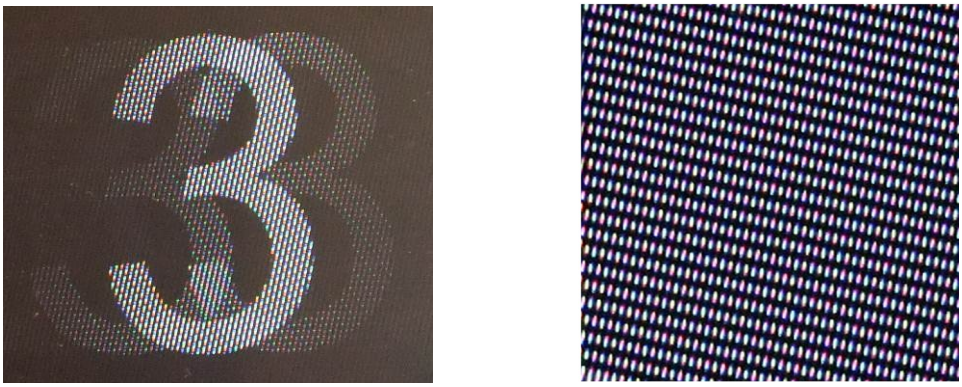
a) **b)** **c)**

Figure 23. Moiré artefacts caused by irregular subsampling: a) test image, b) simulated effect of the optical layers and c) actual photograph of a multiview display, showing the test image in (a).

light diffusing layer which slightly blurs to the image [49], with the aim of decreasing the visibility of Moiré artefacts. Another example of Moiré artefacts exhibited by a multiview display can be seen in Appendix I, Figure 47.

In many autostereoscopic displays, even at the sweet-spot of one view, the contours of one or more other views are still visible. The crosstalk level at the best observation position is known as *minimal crosstalk.* The effect of the minimal crosstalk is especially pronounced in multiview displays where the visibility zones of different views are interspersed and from a given angle multiple views are simultaneously visible [40] [50] [64]. An example image exhibiting multiple ghosting artefacts is shown in Figure 24a. Presence of ghosting artefacts degrade the quality of a 2D image, but are especially damaging for a stereoscopic image. The presence of repeated edges in horizontal direction introduces ambiguity in binocular disparity and can completely destroy the binocular depth cues [61] [15] [65].

In displays with a parallax barrier the barrier creates visible gaps between the pixels as seen in Figure 24b. These gaps are seen as masking artefacts, similar to the fixed-pattern noise



a) **b)**

Figure 24. Distortions in displays with spatial multiplexing: a) multiple ghosts of an image, and b) imaging, or fixed-pattern noise. Reprinted from Displays, Vol 33, no. 2, , A. Boev, R. Bregovic and A. Gotchev, "Visual-quality evaluation methodology for multiview displays,". pp. 103-112. Copyright (2012), with permission from Elsevier.

exhibited by some digital projectors [66]. The perceptibility of masking is limited by physiological factors such as the optical properties of the eye, the density of photoreceptors and the contrast sensitivity function [14]. However, even if separate elements of the mask are visible, the brain has a limited cognitive ability to reconstruct the underlying shape. That ability is known as the *visual Gestalt principle* [3] and the interdependent visibility of patterns with different properties is modelled as *pattern masking* [14].

More information about distortions related to spatial view multiplexing can be found in [60] [52] [63] [64], and also in [P01] [P03] [P05] [P06] and [P08] included in this compound thesis.

3.1.3 Content-related distortions

For a scene on 3D display there is a limited space where an object should appear in order that the object is visible in both eyes. This space is known as *stereoscopic frustum*, and is defined by the positions of the eyes and the size of the display, as shown in Figure 25a. The size of the frustum defines the maximum absolute disparity for objects as a function of their position on the display.

Inside the frustum there is a limited range of disparity values that can be present in stereoscopic content in order for that content to be comfortably observed on a given stereoscopic display. In this thesis such range is called *comfort disparity range*. An example of factors which limit the comfort disparity range is shown in Figure 25b. One of the limitations to comfort disparity range comes from A/C rivalry. Eyes can converge at distances ranging from about 5cm in front of the head to infinity. The eye muscles do not allow the eyes to look in divergent directions. The maximum disparity that can be perceived is limited by the observer's inter-pupillary distance (IPD). If the disparity is larger *divergent parallax* occurs; this is a disturbing, or potentially painful experience [60]. This limitation is somewhat less pronounced in mobile 3D displays, as the mean IPD of 65mm corresponds to substantial part of the display width and the limits imposed by A/C rivalry occur for much lower values.

The subjective experience of a content with excessive disparity is known as *hyperstereopsis* [4], and is considered to be very disturbing artefact, possibly outweighing all other visual artefacts

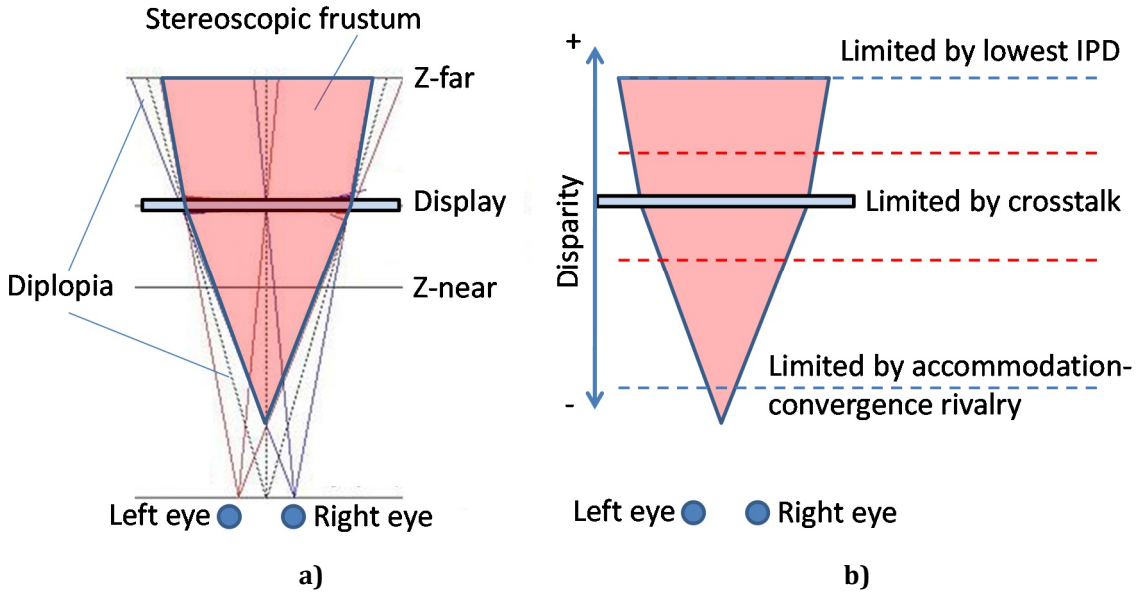


Figure 25. Disparity range of a comfortable perceived content: a) stereosopic frustrum, and b) factors limiting the comfortable disparity range.

in 3D content [15] [61]. By calculating the combined influence of all objective factors, such as A/C rivalry, divergent parallax and visibility of minimal crosstalk, one can attempt to obtain the *objective comfort disparity range*. However, such calculation might not predict the subjective experience well. There are many additional factors that influence the comfort disparity range of a mobile 3D display – such as angular crosstalk, screen reflection, brightness and local contrast of the visualized content and the ability of the observer to locate the sweet-spot of the display. This thesis makes the assumption that there is another, *subjective comfort disparity range*, which is narrower than the objective one and represents the subjective experience of the user and his or her acceptance of 3D content with given disparity.

More information about content-related distortions can be found in [61] [65] [67] and also in [P05] included in this compound thesis.

3.2 Visually important properties of stereoscopic displays

One of the main aspects of this thesis is to address visually important properties of 3D displays. An original idea presented in this work is to relate optical characteristics of 3D displays to visibility of distortions introduced by such displays. This approach allows one to study and measure only these characteristics which are directly related to the subjectively perceived visual quality.

The design of a stereoscopic display is a trade-off between observation convenience and visual quality. There are many optical parameters, such as resolution, brightness, 3D-crosstalk, etc. which influence the quality of such a display. There are number of previous works that deal with estimation of display optical quality, ranging from theoretical considerations regarding the interleaving map [4][5][6] through measuring of the optical parameters of the display [3][7][8] to subjective tests with different multiview displays [9][10][11]. However, using optical parameters of a multiview display to evaluate its quality has two main disadvantages – some parameters, e.g. brightness uniformity are not directly related to the perceived quality; also,

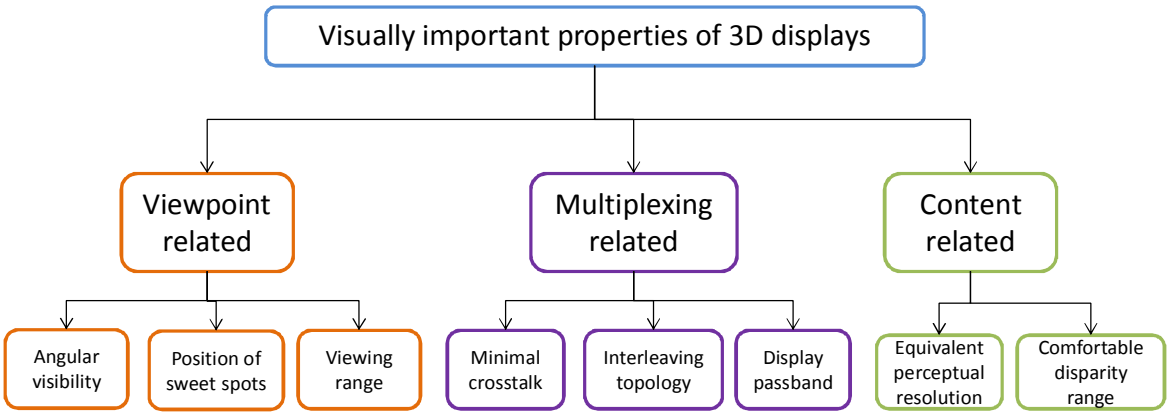


Figure 26. Visually important properties of 3D displays.

visibility of 3D artefacts depends on scene content, observation conditions and HVS properties.

Gaining knowledge of 3D display parameters serves two goals. One is to allow the consumer or content producer to compare the visual quality of two displays or judge if a given 3D content is suitable for a certain display. The other is to use signal processing techniques to mitigate the artefacts in a given 3D display, thus optimizing the visual quality of the output. This section aims to identify the visual characteristics significant from the signal processing point of view and then relate them to visual quality.

Visual parameters of 3D displays are studied according to their potential to create 3D artefacts. By knowing the angular visibility of each sub-pixel in a 3D display one can predict the behaviour of most viewpoint-related distortions, such as crosstalk or colour-bleeding. In order to calculate the influence of A/C rivalry one needs to know the optimal observation position and the viewing range of the display. Knowing the precise position of the sweet-spots can help in estimation of the number of simultaneous observers and potential pseudoscopic regions. Estimation of the comfortable disparity range and the perceived resolution of a given stereoscopic display allows for optimal repurposing of a 3D content.

For predicting the influence of multiplexing related distortions one needs to know the multiplexing topology and the angular visibility of each display element. Knowledge of the topology allows for optimal antialiasing filters to be designed. This dissertation proposes original methods for: 1) deriving the interdigitation topology of a 3D display, 2) measurement of the angular visibility for each TFT element and 3) methodology that combines topology and angular visibility into a *display passband* which quantifies the perceived visual quality of 3D displays. Additionally, the dissertation presents a comparative study on sweet-spot position and comfortable disparity range, measured over a wide range of 3D displays.

3.2.1 Position and size of the sweet spots

In stereoscopic displays the *optimal observation region* is the observation position where a stereoscopic image is perceived with sufficient quality. In passive autostereoscopic displays these regions are small and distinct areas also known as *sweet spots*. However, optimal observation regions also exist in glasses-enabled 3D displays – for example, the crosstalk in a SDPG depends on the observer's elevation. According to [61], 20% crosstalk is a large, but still acceptable for 3D displays and crosstalk beyond 25% is not acceptable. In this thesis, the sweet-

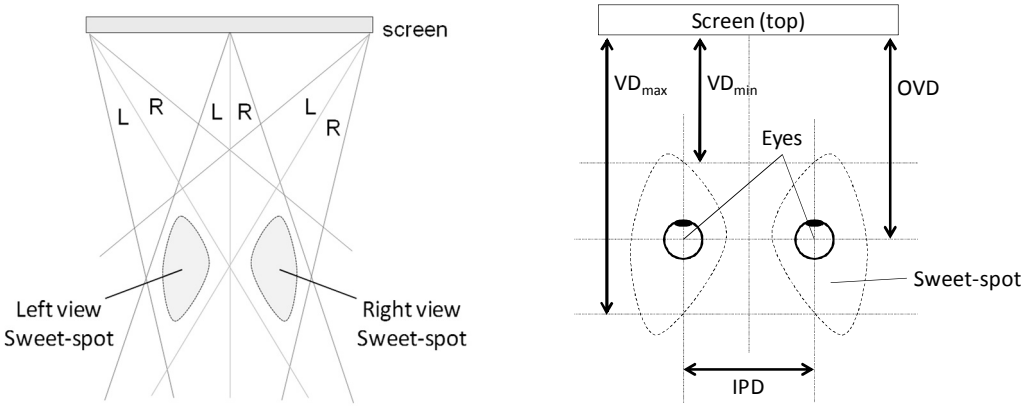


Figure 27. Sweet-spots of an autostereoscopic display: a) left and right sweet-spots, b) optimal, minimal and maximal observation distances. Adapted from [P05].

spot is defined as an observation position where each eye perceives the proper view and the crosstalk between the views is less than 25%.

Since the display is flat, from a given observation position different parts of the screen surface are seen from slightly different observation angles, as shown in Figure 27a. The viewing zone of a view is formed by the union of the visibility zones of each pixel that belongs to that view and has a characteristic diamond-like shape, sometimes referred to as the *viewing diamond* [64]. For comfortable stereopsis both eyes need to be in the corresponding sweet-spot as seen in Figure 27b. This requirement imposes a limit on the range of observation distances suitable for a given display. The size of the sweet-spots can be derived from the angular visibility function, or directly measured using a pair of cameras separated at the chosen IPD. For a given interpupillary distance (IPD) there would be minimum and maximum distance at which both eyes on the observer appear inside the corresponding sweet-spot. These viewing distances are marked in Figure 27b as VD_{max} (*maximum viewing distance*) and VD_{min} (*minimum viewing distance*). For a given IPD there would be an *optimal viewing distance* at which there is an optimal optical separation and lower crosstalk visible across the whole surface of the display.

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	L	L	L	L	L
2	R	R	R	R	R	R
3	L	L	L	L	L	L
4	R	R	R	R	R	R

a)

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	L	L	R	R	R
2	L	L	L	R	R	R
3	L	L	L	R	R	R
4	L	L	L	R	R	R

b)

	Pixel 1			Pixel 2		
	R	G	B	R	G	B
1	L	R	L	R	L	R
2	L	R	L	R	L	R
3	L	R	L	R	L	R
4	L	R	L	R	L	R

c)

Figure 28. Interdigitation maps of dual-view autostereosopic displays: a) row-interleaved, b) column-interleaved at pixel level, c) column interleaved at sub-pixel level.

The optimal viewing distances is labelled as OVD in Figure 27b. Usually OVD, VD_{max} and VD_{min} are calculated using the mean IPD of 65mm.

Naturally, the size and position of the sweet spots is related to the perceived quality. As discussed in [P09], a 3D display with a few, larger sweet spots is considered easier to use than another one with multiple sweet-spots of smaller size. More information about measuring and modelling of 3D display sweet spots can be found in [48] [68] [64] and also [P02] [P05] [P09] included in this compound thesis.

3.2.2 Interdigitation map

The map indicating the relation between the position of a sub-pixel and the view it belongs to is known as *interdigitation map*. Since both TFT-LCD and the optical filter have repetitive structures, the interdigitation map is built from a smaller, repetitive *interdigitation pattern*. The pattern is spatially independent – angular visibility of a sub-pixel depends on its position in respect to the pattern, but not on its absolute position in respect to the display. The interdigitation maps range from simple ones for dual-view displays (see Figure 28) to complex ones for multiview displays (see Figure 29). Most SDPGs have row-interleaved topology as the ones shown in Figure 28a as such topology ensures higher horizontal resolution.

	Pixel 1			Pixel 2			Pixel 3			Pixel 4			Pixel 5			Pixel 6			Pixel 7			Pixel 8		
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
1	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23
2	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1
3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3
4	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5
5	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7
6	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9
7	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11
8	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13
9	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15
10	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17
11	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19
12	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21

Figure 29. Interdigitation pattern of a multiview display.

Autostereoscopic displays have column-interleaved topology since they rely on parallax-based light redirection and views should be separated in horizontal plane. Pixel-based column-interleaving as shown in Figure 28b results in imbalanced colour separation and produces colour-bleeding artefacts. Sub-pixel-based interleaving (shown in Figure 28c) does not suffer from colour-bleeding but the smaller interleaving step limits the size of the viewing zones. Note that autostereoscopic displays can have row-based interleaving as well, provided that the TFT-LCD matrix is rotated at 90 degrees, so its pixel columns appear horizontal [69].

Multiview displays have slanted interdigitation topology where sub-pixels from one view appear along a slanted (in respect to the TFT) line. In order to prevent colour-bleeding the horizontal size of the interdigitation pattern is not divisible by 3, e.g. neighbouring sub-pixels from the same view and on the same row have a different colour, as can be seen in Figure 29. As a result, pixels from one view appear on a non-rectangular grid. In order to design proper sub-sampling filter for that grid one needs to know the precise interleaving topology of the display [48] [52] [55].

In this work an original methodology for deriving the interdigitation patterns of a multiview 3D

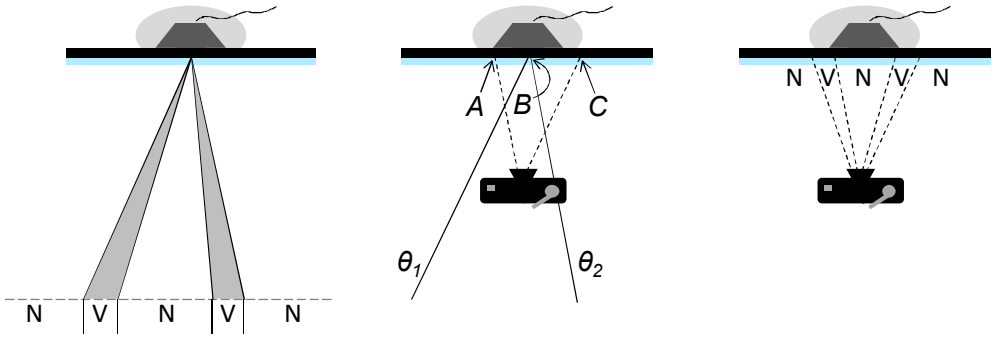


Figure 30. Angular visibility of multiview display: a) visibility separation, b) observation angles when taking a close shot and c) close observation of angular visibility. Appears in [P04], © 2010 Society for Information Display, reprinted with permission.

display is proposed. Ideally the view should be visible with full brightness from a limited range of observation angles (marked with "V" on Figure 30a) and be invisible from anywhere else (as marked with "N" in the same figure). A group of sub-pixels with similar angular visibility have a higher N/V ratio than a group of sub-pixels with varying optimal observation vectors.

The straightforward method for finding sub-pixels which belong to one view is to turn on a group of sub-pixels and to study the angular visibility of the resulting image. One way to do it is to photograph the display from closer than the optimal observation distance, as shown in Figure 30b. In the proposed method such photograph is referred to as a *close shot*. Following the assumption for spatial independence of angular visibility, the visibility points along the horizontal axis would correspond to the visibility of one point as seen from different angles. As exemplified in Figure 30b, point "A" as seen from the camera should be the same as the visibility of point B as seen from observation angle θ_2, and point "C" as seen from the camera should be the same as the visibility of point B as seen from observation angle θ_1. In the close shot the ratio between visible and invisible parts is proportional to the N/V ratio of the pixel group under test, as shown in Figure 30c. The group of sub-pixels with the highest N/V ratio belongs to the same view. A close shot of rows of sub-pixels with various N/V ratio can be seen in Appendix I, Figure

49.More details about the procedure for finding the interdigitation topology of a 3D display can be found in [P04], included in this compound thesis.

3.2.3 Angular visibility

Knowing the angular visibility function of each display element allows one to predict the position of the sweet-spots and the crosstalk for different observation positions. Measuring the brightness of a single pixel by photographing the display would be a tedious and noise-prone task. Instead, since sub-pixels in one view are supposed to have the same angular visibility, one could measure the mean brightness of a view and assign it to each pixel of that view. Another problem is measuring the view brightness as a function of the angle – inaccurate camera placement would result in the angular visibility function being sampled at irregular intervals. This can be solved by measuring the visibility of each view at some selected points and search for single function that gives the best fit for all measurements regardless of the angle.

This thesis introduces a methodology for measuring the angular visibility without the need of precise measurement equipment. The first step is to prepare two groups of test images. The first

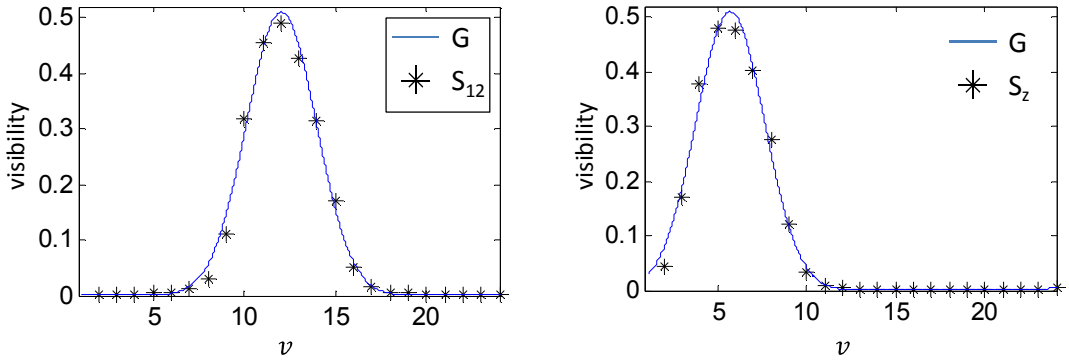


Figure 31. Modeling angular visibility for a multiview display: a) shape of derived Gaussian curve (G) with measurements for one observation point (S_{12}), b) predicted visibility of all elements between two observation points, vs. with measurements done in that location (S_z). Appears in [P04], reprinted with permission.

group consists of so-called *single view* images, where only the sub-pixels from one view are lit. These images are used for measuring the angular visibility. The second group contains test images where all pixels are set to different levels of grey in order to linearize the camera response function [70]. Each test image is shown on the test display and is photographed from a number of observation positions. The observation positions are selected on a line parallel to the display surface and at the optimal viewing distance. If the measurement point is displaced from the center of a visibility zone, the visibility function gets sampled with an offset and the maximum value of that function falls in between two samples. However, judging by measurement results in the literature [64] [68] [71] one can assume that the visibility function for all observation points can be closely approximated by the same function, which has its peak occurring in the optimal observation spot for the corresponding view. Based on this assumption, one can search for single function that closely fits measurements for all positions regardless of possible offset. An example of such a function is shown in Figure 31a. Once the function is derived, it can be used to approximate the visibility of each element for an arbitrary observation angle, as shown in Figure 31b.

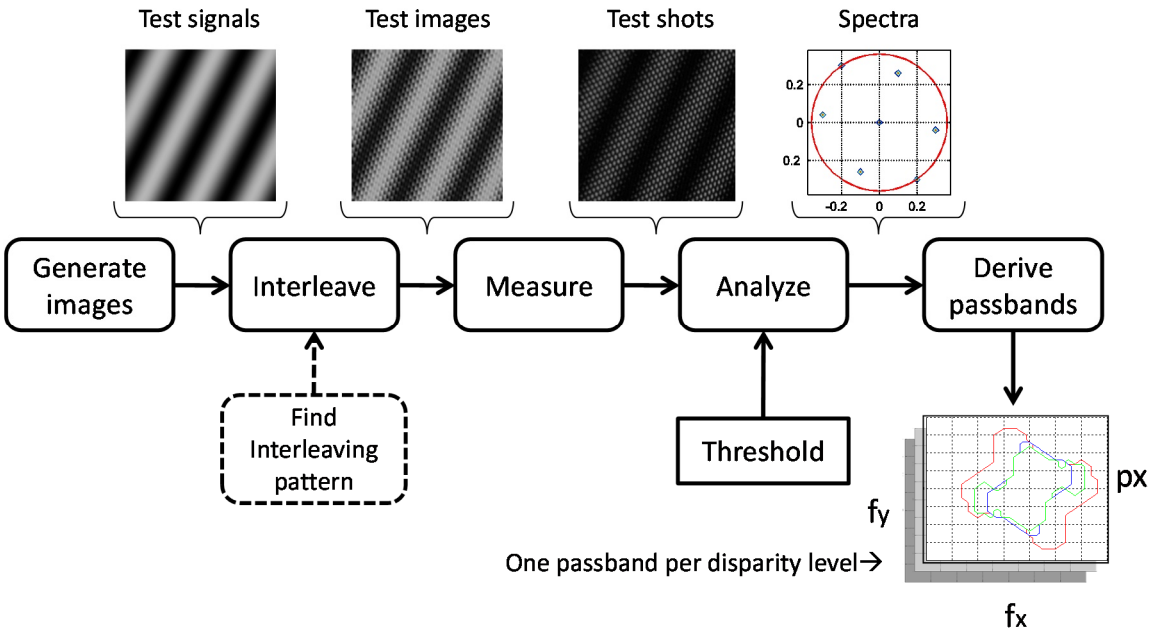


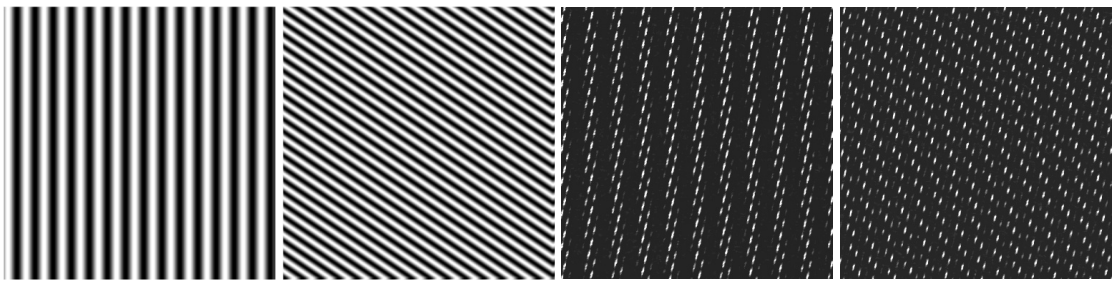
Figure 32. Block diagram of the methodology for deriving the display passband. Adapted from [P01].

More details about measuring the angular visibility function of a 3D display can be found in [48] [64] [68] [72], and also in [P04] [P11] included in this compound thesis.

3.2.4 Display passband

Spatially-multiplexed 3D displays suffer from masking distortions and fixed-pattern noise caused by visible gaps between the pixels or by apparent non-rectangular shape of a pixel. The visibility of such distortions depends on the frequency components of the visualised content. In order to assess the visibility of masking one needs to study the performance of the display in the frequency domain. A methodology for finding the passband of a spatially multiplexed 3D display is proposed in this dissertation. It contains five steps, as shown in Figure 32.

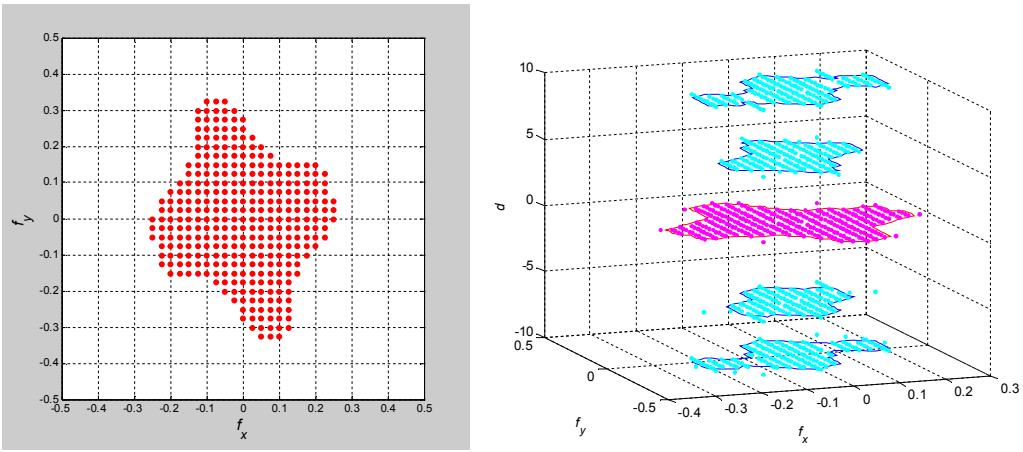
The first step is to prepare number of *test signals* which contain a 2D sinusoidal pattern with varying horizontal and vertical frequency components, as the ones shown in Figure 33a and Figure 33b. Then each test signal is used to prepare a number of *test images* with different



a) **b)** **c)** **d)**

Figure 33, Deriving the passband: a) test image with horizontal frequency component, b) test image with horizontal and vertical frequcency component, c) observation of the first test image, where the intended frequecy is still tominant, d) observation of the second test image, where the intended frequency is masked by imaging distortions.

apparent depth. Apparent depth is created by mapping the same signal to each view of the display, introducing different amount of disparity to each view and interleaving all the views. The third step involves automated visualization of all test images on the display and making a snapshot of each one with a high resolution camera. The output of that step is a collection of *test shots*, similar to the ones shown in Figure 33c and Figure 33d. In the next step the *spectrum* of each test shot is analysed and the amplitudes of two components are measured. The component is created by the *intended signal*. It appears as a peak on the same place in the spectrums of the test signal and the test shot. The position of that peak also determines the *circle of interest* which is centred at the point of origin and passes through the intended signal peak. The second component in the measurements is the *most visible distortion*. Display distortions introduce multiple peaks in the frequency domain. The largest peak which appears inside the circle of interest is marked as the most visible distortion component. The ratio between the intended signal amplitude and the most visible distortion amplitude is used to derive the *display passband*.



a) b)

Figure 34. Passband of a 3D display: a) samples inside the passband; b) passbands for different disparities. (a) appears in [P04], reprinted with permission from Elsevier, (b) appears in [P01], reprinted with permission from the Society for Information Display.

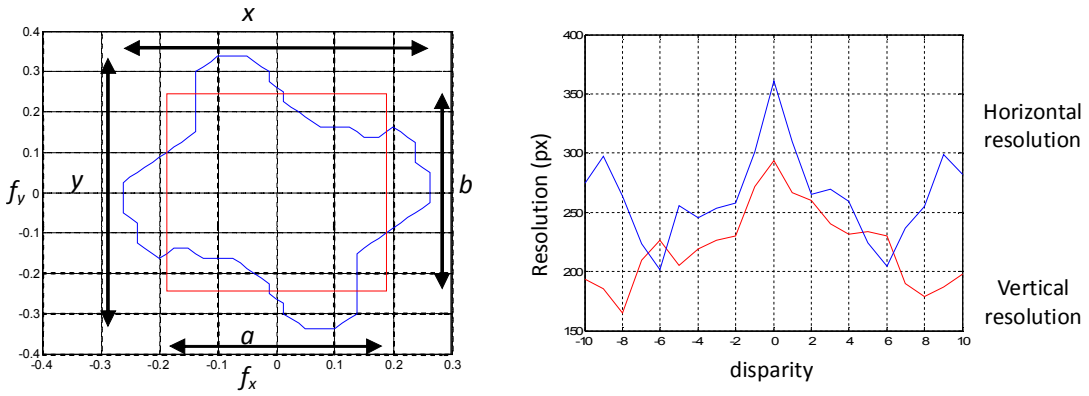
Test shots, for which this ratio is smaller than a threshold, are marked as being inside of the passband and all other test shots, as being outside. For example, the test pair shown in Figure 33a (test signal) and Figure 33c (test shot) belong to the display passband, as the intended signal frequency is dominant in the test shot. As the HVS can reconstruct missing elements of a structure, the vertical bars in Figure 33a are still visible in Figure 33c. The test pair shown in Figure 33b and Figure 33d does not belong to the passband since the dominant frequency in the test image is masked by the imaging distortion. By scanning the frequency domain using multiple test images one can sample the passband, as shown in Figure 34a. In step 5 all frequency components which passed the threshold are combined into the display passband area. The passband area represents the ability of the display to faithfully reproduce image signals with spatial frequencies within a given area. The output is a measurement for the display passband for test signals with given disparity. Finally, the all passbands measured for different disparities are collected into a 3D passband area, as shown in Figure 34b. The shape and the size of the display allow for quality comparison between 3D displays. A display with a larger and more uniform passband is assumed to be of higher visual quality as it can faithfully represent

larger range of image details. Additionally, by knowing the frequency characteristics of a 3D scene, content producers can judge if the scene would "fit" the passband of a given display, resulting in a faithful representation. More details about deriving the passband of a 3D display can be found in [P01] [P04] included in this compound thesis.

3.2.5 Equivalent perceptual resolution

Although the display passband allows comparison of quality between displays, it is not straightforward to use it for judging the quality of a single display. However, most display users have an intuitive idea about the image quality of a display with a given resolution. This dissertation proposes a method of calculating the equivalent perceptual resolution of a spatially-multiplexed 3D display.

In order to relate the 3D passband to a 2D display resolution, one can approximate the passband for each disparity with a rectangular shape. The main idea is to have a rectangle centred at origin that will have the same area (in size) as the original passband and overlapping as many passband points as possible. Another requirement is to keep the aspect ratio between the



a) b)

Figure 35. Equivallent resolution of a 3D display: a) approximation of the passband with a rectangle; b) equivallent resolution for different disparities. Adapted from [P01].

maximum values in horizontal and vertical directions. By observing these two constrains (area and aspect ratio) one can find the rectangle which is a "best fit" to a given passband area, as shown in Figure 35a.

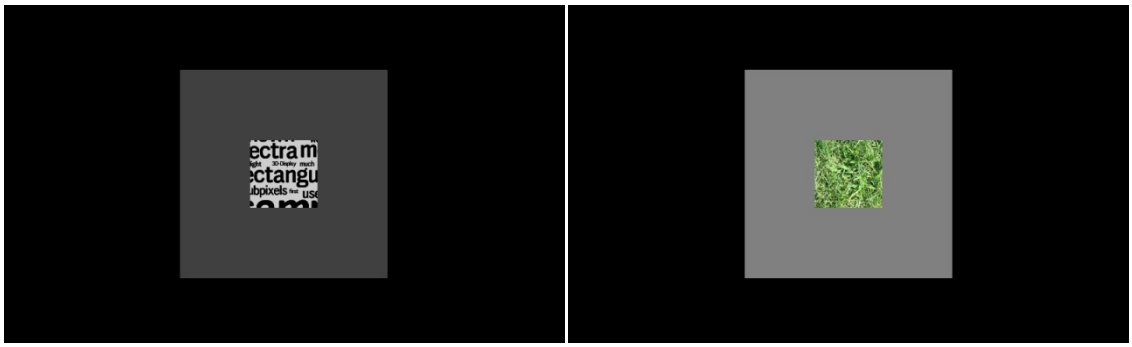
In order to represent this figure in a more understandable way, one can convert rectangular passband sizes to equivalent resolution in pixels. This is done by multiplying the passband width (or height) by the overall resolution in horizontal (or vertical) direction. An example for equivalent resolution derived for a 24-view 3D display, plotted as a function of the object disparity, can be seen in Figure 35b. Notably, the function is not monotonic but has local maximums for some disparities. Knowing the equivalent resolution of a 3D display for different disparities can help content producers to rearrange placement of objects in a 3D scene in order that each object is seen with optimal quality.

3.2.6 Comfortable disparity range

There are a number of parameters which determine the maximum disparity range which can be comfortably observed on a 3D display. Some of them, like divergent parallax, A/C rivalry and

frame violation can be calculated, provided that one knows the display resolution, pixel density, observation distance and IPD of the observer. However, other (and less studied) parameters are probably involved as well – for example subjectively perceived contrast, screen reflection index, room illumination, etc. The unambiguous way to determine the comfortable disparity range of a 3D display is to perform subjective tests where the acceptance of 3D content is rated. Naturally, the main variable in the experiment is disparity range. Since local contrast of the content greatly influences the perceptibility of ghosting artefacts [15] [60] [61], the content under study should contain scenes with various levels of contrast. As contrast perception is frequency dependant [8], acceptance of 3D content is possibly affected by frequency characteristics of the image.

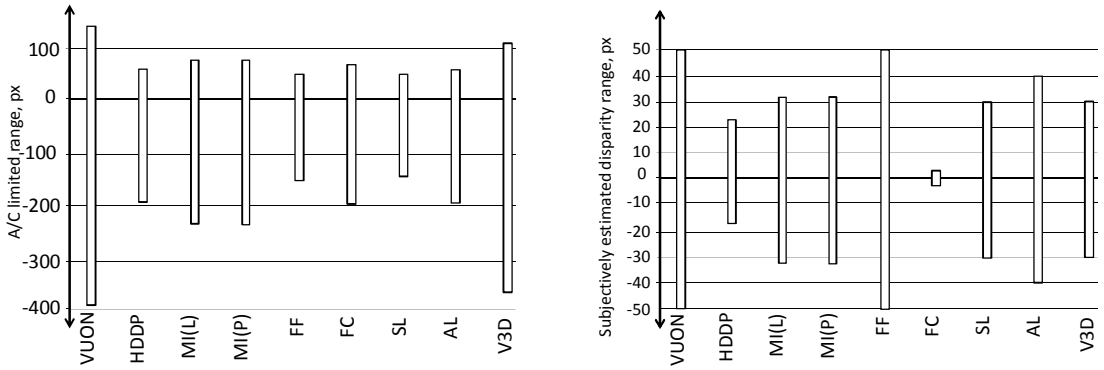
One of the goals of this dissertation was to characterize and compare comfort disparity range of different 3D displays. In order to assess subjective influence of disparity range over 3D quality a small-scale subjective experiment was designed and carried out. A group of 10 observers was asked to rate the acceptance of a number of test images. The images contain two patterns, as shown in Figure 36. One pattern contains text (sharp image with high contract), and the other has natural content (smooth image with low contrast). With each pattern a number of images with varying local contrast are created. The contrast is altered by changing the brightness of the patch and that of the background. Finally, each test image is used to generate a number of stereoscopic pairs with varying disparity. Observers were asked to rate each stereoscopic pair. The test was repeated for 9 different 3D displays.



a) **b)**

Figure 36. Test images, used for estimation of the subjective disparity comfort zone: a) high contrast image, b) low contrast image.

The comfort disparity range for each display was calculated using objective parameters. The group of ranges is shown in Figure 37a. Based on the test results the subjective disparity range was derived for each display, as shown in Figure 37b. From the figures one can see that the subjective comfort disparity range is 4 to 5 times smaller than the objectively calculated one. Apparently, the influence of display properties influences the range more than the viewpoint-related parameters. More information about comparing various parameters of 3D display can be found in [50] [43] and in [P05] included in this compound thesis.



a) **b)**

Figure 37. Comfortable disparity range of a various 3D displays: a) calculated using objective parameters, b) derived from a subjective test.

4 Visual optimization

One of the major goals of this dissertation is to develop a set of signal processing algorithms which allow 3D content to be visually optimized for a given spatially-interleaved 3D display. This section proposes a number of original algorithms for 1) view-point optimization, which use observer-tracking and adapt the content in real-time for the position of the observer, 2) passband optimization which filters the scene texture to match the throughput of a given display and 3) content optimization which adjusts the scene geometry according to the crosstalk and comfort disparity zone of a 3D display.

Signal processing techniques can be used for improving the visual quality in three ways. If a distortion introduced by the display can be described as an invertible function, one can pre-process the image using the inverse function. In such case the changes caused by pre-processing would undo display distortions, resulting in a clean signal representation without artefacts. This process is known as *pre-compensation* and can be used to improve some cases of pseudoscopy, hyperstereopsis and ghosting. In case of distortions which cannot be pre-compensated, a signal

Table 1 – Visual optimization: distortions, artefacts and mitigation algorithms

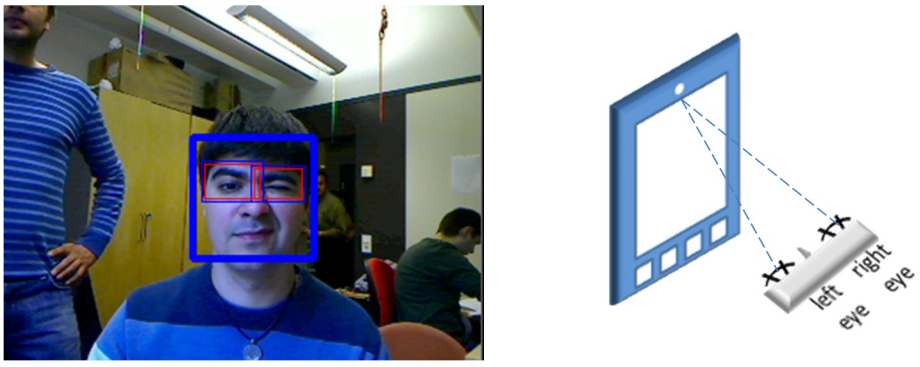
Distortion source	Artefact type	Artefact mitigation algorithm	
		Dual-view displays	Multiview displays
Observation angle	Pseudoscopy, ghosting	Pseudoscopy correction	Extended head parallax, extended viewing distance
Crosstalk	Ghosting	Pre-compensation	Crosstalk mitigation
Aliasing	Moire	Antialiasing filters (1D)	Antialiasing filters (2D)
Passband	Moire+ghosting+masking	Pass-band optimization (2D/3D)	
Excessive disparity	Hyperstereopsis	Content re-purposing	

processing algorithm can decrease their visibility, thus helping mitigate the perceived annoyance of artefacts and improving the quality. Artefact mitigation algorithms are possible for imaging, aliasing and cases of pronounced crosstalk. Finally, the visibility of some artefacts does not depend purely on the content, but also on observer position, motion and head orientation. Such cases need real-time algorithms which actively track the observer and process the visual signal accordingly.

A list of artefacts and appropriate artefact mitigation techniques can be seen in Table 1. In order to mitigate distortions caused by observation angle one needs to know the position of observer in respect to the display. Generally this is achieved using camera-based tracking and face- or eye-tracking algorithms. Once the observation position is known, the image can be optimized for the calculated angle and distance. Algorithms for viewpoint optimization usually work for one observer only; however there are solutions which allow 3D display to adapt to the observation position of multiple observers simultaneously [43] [46] [73].

Ghosting artefacts can be either pre-compensated or mitigated. For dual-view displays, where crosstalk levels are low, pre-compensation is possible, but limits the dynamic range of the display [74]. Crosstalk pre-compensation is possible both for time-sequential and spatially-

multiplexed dual-view 3D displays. A similar approach can be used for a multiview display if a single observer is tracked. However, the possibility of multiple observers, and the pronounced crosstalk between neighbouring views, make crosstalk mitigation the preferred approach for multiview 3D displays. Such algorithms aim to reduce the visibility of ghost images by filtering horizontal high-frequency components of the image at the expense of losing image details. The range of artefacts, which are caused by the optical separation layer of a multiview display can be mitigated by antialiasing filters [52] [53] [55], or by deriving the passband of the display and preparing a filter with removes image data with frequency components outside of the passband [P03] [P06] [P07]. Such passband filtering can be done either by a single 2-D filter (as proposed in [P07]), or by a bank of 2-D filters, each one optimized for a different image disparity (as proposed in [P03] [P06]). If the scene is represented as an epipolar volume, one can implement pass-band optimization as a 3-D filter [53]. Finally, excessive disparity can be compensated by a transformation which alters the disparity range of a scene. Such transformation can be a combination of image rescaling and cropping, or, if more processing power is available, a combination of dense depth estimation and image warping algorithms can be used [75].



a) **b)**

Figure 38. Head and eye-tracking: a) face and eye tracking b) position of the user in respect to the front facing camera. (b) appears in [P12], published by EURASIP.

4.1 View-point optimization

In order adapt the display to the observation position of the user, an artefact mitigation algorithm should detect and track the position of observer's eyes. The eye-tracking should work in real-time because a tracking delay might optimize the image for wrong observation position and introduce visible artefacts.

User-tracking algorithms for 3D displays are usually designed for a single observer only. In dual-view 3D displays user-tracking is used to avoid pseudoscopy. In displays with passive optics this is done by re-arranging the content and flipping the channels. One algorithm which can address both position and pose of observers head is proposed in this thesis. In dual-view displays with active optics (for example, Free2C [76]) user-tracking is used to dynamically adapt their sweet-spots to the position of the viewer. In multiview 3D displays user-tracking is used to avoid the stereo-edge. One algorithm that re-addresses the content on a multiview display and provides extended head parallax is introduced in this dissertation. However, solutions which involve tracking of multiple users do exist. Toshiba announced a 3D TV model which uses a combination of a multiview display with passive optics, and tracking of multiple users [46]. A hybrid approach which combines active optics and multiuser head tracker has

been under development [43] [77]. Such a combination theoretically allows a display to provide an extended head parallax provided to a number of observers.

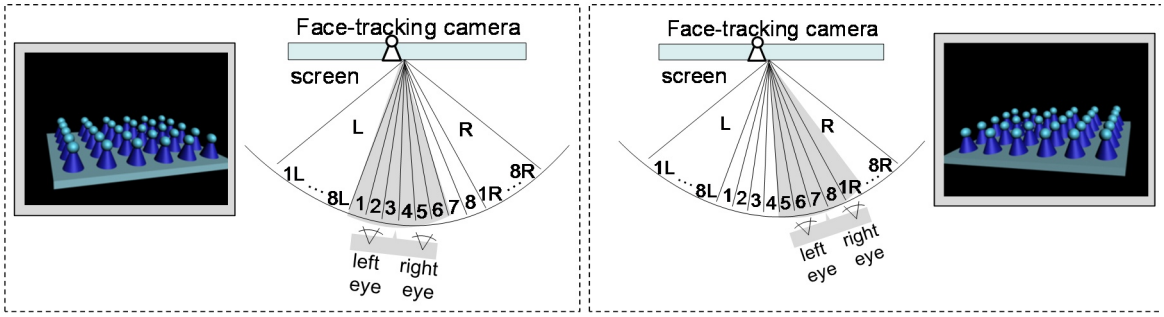
Multiuser observer-tracking algorithms have been discussed in [78] (using head-tracking) and [79] (using eye-tracking). A real-time face- and eye tracking algorithm working on a mobile platform is presented in [P12]. The implementation allows splitting the processes of face and eye detection between the ARM and DSP cores of an OMAP 3430. In order to increase the face detection speed, the algorithm searches for a subset of all possible face sizes as the visible face size is limited by the requirement to the user to stay within the visual comfort zone. Face detection is performed by a two-stage hybrid algorithm which combines skin detection with feature-based face detection [80]. It is implemented on the ARM core. If a face is present, eye-detection is performed only in the top half of the detected region, as shown in Figure 38a. The eye detection is implemented on the DSP core. It detects the eyes using a Bayesian classifier working on Dual-Tree Complex Wavelet Transform (DT-CWT) features [81] [82]. The combination of both algorithms allows precise detection of the position of the eyes in respect to the camera, as shown in Figure 38b.

4.1.1 Optimization for observation angle

Visual optimization for observation angle is solved differently for dual-view and for multiview 3D displays. In dual-view displays the most pronounced viewpoint related distortions are pseudoscopy and ghosting. Ghosting artefacts are seen if any of observer's eyes appears inside of the stereo-edge (between the visibility zones of two views). Pseudoscopy is experienced if both eyes appear in the visibility zones of the opposite view. In all other cases, both eyes appear in the visibility zone of the same view and 2D image is perceived. One interesting feature of dual-view autostereoscopic displays is that some models allow switching between 2D and 3D modes which enables the display to "fall back" to 2D image and regain display resolution.

An algorithm for observation angle-based optimization for dual-view 3D display is proposed in [P12]. Based on the horizontal coordinate of the pupil, three tracking zones are defined: visibility zone of the left view (marked with "L" on Figure 22b), visibility zone of the right view (marked with "R" on the same figure) and zone with high crosstalk (marked with "X"). Pseudoscopy is avoided by flipping the left and right channel if an eye is detected to be in the opposite viewing zone. Ghosting artefacts are avoided by switching the parallax barrier off and switching the content to "2D" if any of the observer's eyes appears inside of an "X" area. The rationale for this rule is that if one eye of the observer perceives excessive crosstalk, stereoscopic perception is not possible, and it is preferable that the observer does not see the ghost artefacts as well.

In multiview displays, the observation zones of neighbouring views are interspersed and ghosting artefacts can hardly be compensated in real-time. In these displays, the most pronounced view-point artefact is the limited area, where head-parallax is visible and the severe ghosting visible at the edges of that area. A "semi-active" approach for extending head parallax and removing the ghosting in the stereo-edge is proposed in [P09]. It combines the precise light redirection of a multiview display, a single camera and less-precise sub-real-time head-tracking. The software part of the system that takes care that the observer's head is "surrounded" by a group of properly rendered views. Once the approximate position of the observer's head is found, the precise delivery of different images to the eyes is handled by the (passive) multiview optics. As shown in Figure 15c, each view is seen from a number of observation spots and the whole set of zones is repeated on the sides. For an observer moving



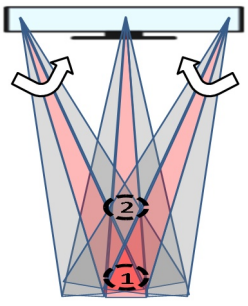
a) b)

Figure 39. Selective view updating for continuous parallax: active views and visualized scene percepcive for a) one observer position, b) another observer postion. Adapted from [P09].

laterally in front of the display, the stereo-edge between the first and the last view produces a characteristic "jump" of the 3D image [71]. However, one can provide a continuous parallax by replacing the views which are not visible with observations of the same 3D scene from new angles. For example, when the user's head is positioned as seen in Figure 39a, the active views are from 1 to 6, and views 1 and 5 are seen by the left and right eyes correspondingly. When the user moves to the position shown in Figure 39b, views 5 and 6 show the 3D scene at the same angles as before and view 1 is updated to show the scene at a new angle. In reality, the eyes of the user fall into neighbouring views and the view update happens well outside of the eye position. The head-tracking has only to ensure the head of the observer is approximately at the centre of the set of updated views. Unlike the "active" eye-tracking approach, estimation of the distance between the observer and the display is not needed as a set of properly rendered views can provide proper parallax to the eyes in a wide range of head positions. Also, real-time performance of the system is not necessarily critical as the user is always "surrounded" by a safe margin of properly rendered views.

4.1.2 Optimization for viewing distance

Both dual-view and multiview autostereoscopic displays are designed to be viewed at a particular distance. At the optimal viewing distance the intended view is seen across the whole surface of the display, as marked with "1" on Figure 40a. At a distance closer than the optimal the observer sees different visibility zones at the left and right edges of the display, as marked with "2" on the same figure. If the distance to the observer is known, the content on the display



2	1	0	0	0	0
2	1	0	0	0	-1
2	1	0	0	0	-1
2	1	0	0	-1	-2
1	0	0	0	-1	-2
1	0	0	0	-1	-2

F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F
F	X	0	0	X	F

a) b) c)

Figure 40. Distance-based content optimization: a) re-routing of views for observation distance, shorter than the optimal, b) example re-routing table for stereoscopic display, c) example rerouting table for stereoscopic display.

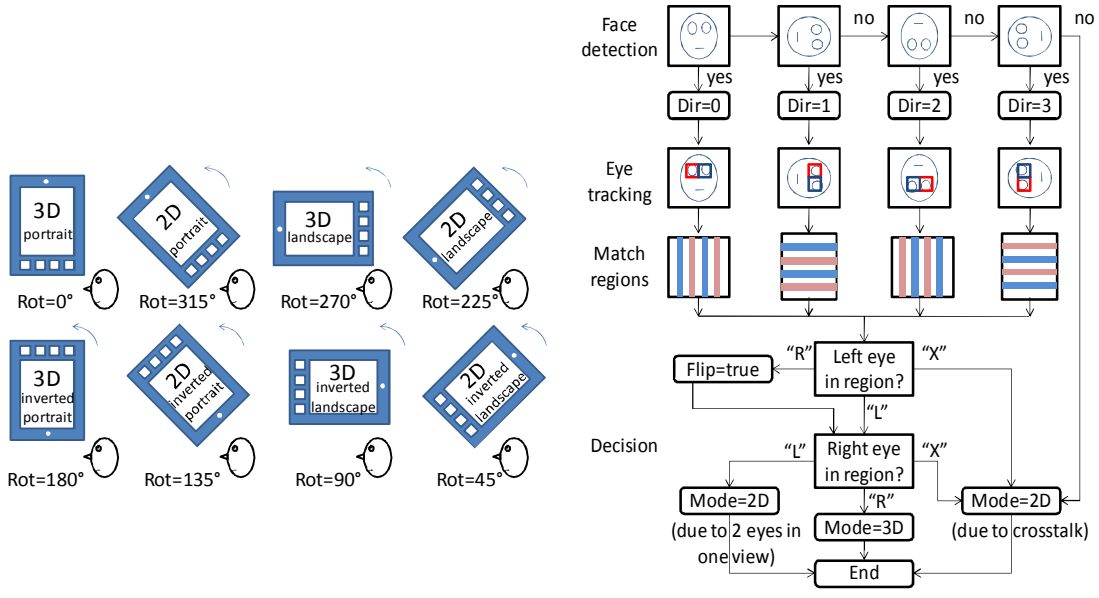
can be re-rendered accordingly. In order to measure the distance to the observer, eye and face-tracking is performed by two cameras simultaneously. For more information on the algorithm, the reader can refer to [83].

In the case of multiview display the information is shifted between the views – for example, the image along the right edge of the display intended for the central view (marked with red on the Figure 40a) can be rendered in the previous view (as shown by the curved arrow). The opposite is done along the left edge. This procedure can be expressed as a re-routing table which optimizes the image for a given observation distance. The re-routing table should be re-calculated for any given distance to the observer. In the case of a multiview display, pixels intended for certain view would be re-routed to other views. An example of a multiview rerouting table is given in Figure 40b. The surface of the display is separated into sub-sections, and the number in each subsection indicates the rerouting operation to be performed in the corresponding area of the display.

In the case of stereoscopic displays the re-routing table looks like the one given in Figure 40c. In this table "0" means that the pixels in the corresponding area are left unaltered. The pixels in the "F" areas should be "flipped", effectively swapping the pixels intended for the left and right view. The areas marked with "X" would be perceived with excessive crosstalk because for these areas the observer appears between the viewing zones of the left and right views. In the "X" areas, a monoscopic image should be projected by copying all pixels from one view to the other.

4.1.3 Optimization for observation pose

Some dual-view autostereoscopic 3D displays with a parallax barrier have the ability to switch between horizontal 3D and vertical 3D modes. They can provide 3D effect both in landscape and portrait orientation. For such displays a visual optimization algorithm can select the 3D mode and scene orientation based on the orientation of the observer's eyes as illustrated in Figure 41a. For observation angles at which 3D effect is not possible, the system switches the display



a) b)

Figure 41. Selection of display mode and scene orientation, according to the orientation of the eyes of the observer: a) mode and orientation according to rotation angle, b) block diagramm of the algorithm. Appears in [P12], published by EURASIP.

into 2D mode. An eye-tracking algorithm for selecting scene orientation is proposed in [P12]. The block diagram of that algorithm is shown in Figure 41b. First, face detection is attempted four times, each time rotating the input image by 90 degrees. If detection fails, the presumption is that either the face of the observer is too far from the display or it is at a wrong angle. In both cases 3D perception is not possible and the system switches the display into 2D mode. If face detection is successful, eye-tracking is performed according to the orientation of the detected face. The position of the eyes is matched against the map of observation zones of each view (see also Figure 22b). If both eyes are found in the corresponding regions, the system switches into 3D mode. If both eyes appear in the regions of the opposite view the system flips the channels and switched into 3D mode. If both eyes fall into the observation zone of the same view, or at least one eye falls in an inter-zone crosstalk area, the system switches into 2D mode.

4.2 Display passband optimization

When visualizing images on spatially-multiplexed displays there are two potential sources of distortions; aliasing due to the picking up of sub-pixels on non-rectangular grid and imaging, due to the presence of gaps. While various antialiasing filters for 3D displays have been proposed, pre-filters addressing imaging artefacts of 3D display are a novel idea introduced in this thesis.

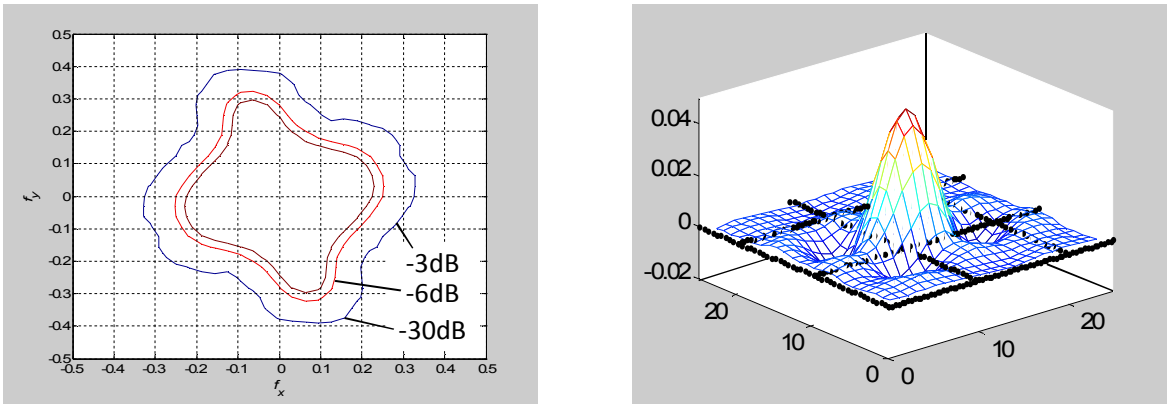
In [63] Jain and Konrad introduced a method for designing 2D non-separable antialiasing filters for an arbitrary sub-sampling pattern. They devised a 2D filter with a passband that spans all frequencies at which the contribution of all alias terms is smaller than the original signal itself. In [55], Moller and Travis used simplified optical filter model to analyse display bandwidth and derived a spatially-varying 2D filter which requires knowledge of scene per-pixel depth. In [53] Zwicker *et al*. proposed a low-pass filter to be applied on the sampling grid of the multiview display expressed in ray-space which aims at preventing both intra- and inter-perspective aliasing. However, their model does not take into account the directionally dependant aliasing caused by the slanted optical filter.

Usually, imaging is tackled by an anti-imaging post-filter. As the imaging is created by the physical structure of the display, it is impossible to impose a post-filter. However, the effect can be partially mitigated by a pre-filter. In order to determine the properties of the required 2D filter, and consequently have the best possible representation of images on the display (minimizing aliasing, imaging and ghosting), it is necessary to determine the performance of the display in the frequency domain; that is, one has to know which frequency components in the image can be kept (ones that will be properly represented on the screen), and which ones have to be attenuated as potential causes of distortions.

The method proposed in this dissertation aims to derive the frequency response of the display, and devise a filter which removes image components which would cause visible distortions. The region containing frequencies that are properly represented on the screen is the passband of the display, and all other regions are its stopband. In order to improve the image quality, one should design a filter which mitigates frequency components in the stopband. Such a filter would address both aliasing and imaging artefacts.

4.2.1 Passband approximation with a non-separable filter

The design of non-separable passband-optimizing filter is discussed in [P03] and [P07]. As a practical example, such a filter is designed for a 24-view 3D display, which has passband as the one in Figure 34a. For that display, the shape of an ideal 2D antialiasing filter is as shown in



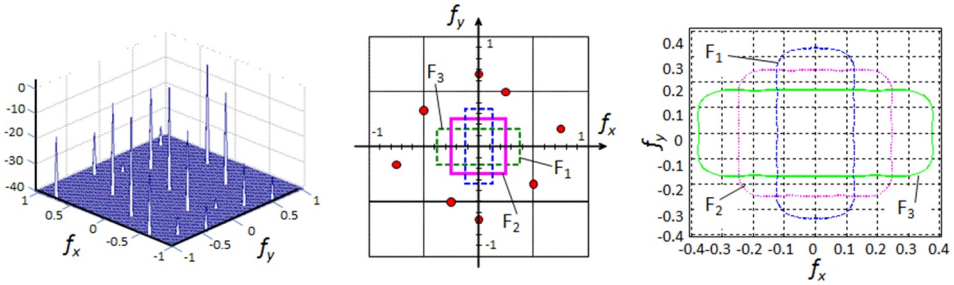
a) b)

Figure 42. 2D non-separable filter: a) magnitude response – contour plot for -3, -6, and -30 dB, b) impulse response. Appears in [P04], © 2010 IET, reproduced with permission.

Figure 42a. In this figure, the curve shows the ideal cut-off frequency, that is, the passband of the filter should be inside the contour, and its stopband everywhere else. For designing a non-separable 2D filter approximating this ideal one, the windowing design technique with the Kaiser window of length 24 has been used [84]. The designed filter is 2D non-separable, with size of 24 by 24 and impulse response as shown in Figure 42a. The corresponding magnitude response (contour) of the designed filter is shown in Figure 42b. The Kaiser window has been selected as a good candidate due to its narrow transition band and flexible attenuation. The variable parameter of the Kaiser window controlling the stopband attenuation has been set to β =2.2. Such selection will ensure a stopband attenuation of at least 30dB that is good enough for the display under consideration. The -6dB line in Figure 42a approximates the ideal cut-off frequency. Due to the finite transition bandwidth of the designed filter, even after applying it to the input image, some aliasing errors will occur on the display. However the aliased frequencies will be attenuated by the filter (either filter transition band or stopband) and as such they will not be visible. The filter size of 24 by 24 has been chosen as a good compromise between the implementation complexity, transition bandwidth and approximation of the ideal filter.

4.2.2 Passband approximation with a separable filter

The computational complexity of a 2D filter is rather high. Considerable computational savings are achieved if the 2D filter can be separated into two 1D filters, one filtering in the horizontal



a) b) c)

Figure 43. 2D separable filter: a) spectrum of sub-sampling pattern for one view, b) possible solutions for optimal antialiasing filters, c) magnitude responses of these filters, -6dB contour plots. Adapted from [P03].

direction and one in the vertical direction. The design of separable passband-optimizing filter is discussed in [P03] and [P07].

By knowing the interdigitation pattern and angular visibility of each element one can derive the pattern of visible pixels as seen from the sweet-spot of one view. As discussed in Section 2.2.5, this pattern behaves as sub-sampling mask. As an example, the mask spectrum of a 24-view 3D display is shown in Figure 43a. Each of the peaks in this spectrum corresponds to a source of aliasing. In order to avoid Moiré artefacts, a filter has to be designed in such a way that its passband does not overlap with any of its copies generated by moving its centre to any of those aliasing sources. It is possible that there are several different separable filters that can be used as antialiasing filters for this display, as shown in Figure 43b. Each of those filters will perform proper antialiasing, but due to different shapes the visual quality of displayed images will be different. Which separable filter would yield best visual results depends on the content. The experiments presented in [P07] suggest that for textual information such as subtitles, filters whose passband is close to square perform better than filters with elongated passbands. For designing 1D filters with the desired cut-off frequencies, the windowing technique with the Kaiser window of length 24 can be used. As an example, the magnitude responses (contour) of the separable 2D filters optimal for the said 24-view display are shown in Figure 43c.

4.2.3 Passband approximation with a tuneable filter

The results in [P07] suggest that the filter that fully suppresses aliasing does not always give the best perceptual quality. Some people prefer sharper-looking images at the expense of some Moiré artefacts. In order to allow the user to control the antialiasing process according to his/her own preferences, one can design a set of tuneable filters which depend on two parameters – apparent depth and desired sharpness. The sharpness parameter is expressed in terms of signal-to-distortion ratio; this is expected to affect the visibility of aliasing in perceptually linear fashion, regardless of the apparent depth.

An artefact mitigation framework using tuneable filters is proposed in [P06]. It allows the user to specify the percentage of visible distortion over the original signal. The framework does the

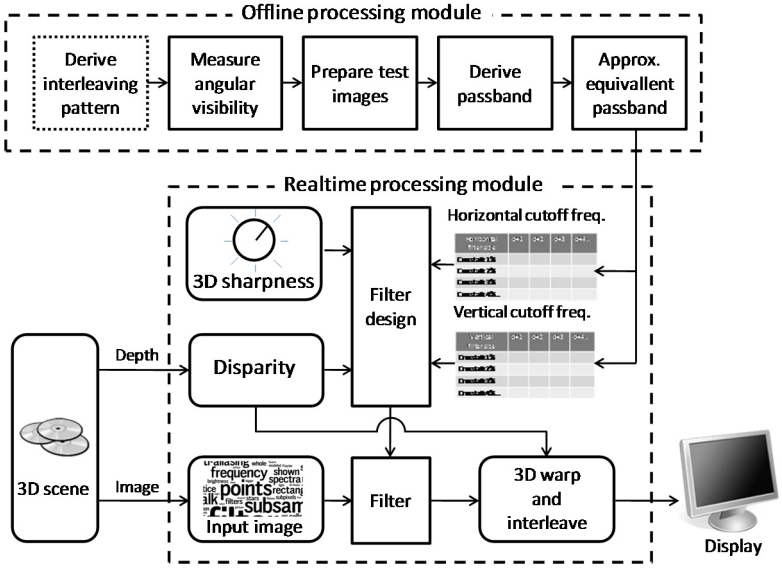


Figure 44. Passband-optimization framework with tunable "3D sharpness" level. Appears in [P06], © 2011 SPIE, reprinted with permission.

necessary processing to maintain the distortions within the selected limit, taking into account the display passband for different disparity values. It consists of two modules as shown in Figure 44. The first module consists of off-line display measurements. The second module performs real-time image filtering, in accordance with scene content and user preferences. During the offline measurements the passband of the display is derived for a range of disparity values. Each passband is approximated by a rectangle. The resulting passband areas are stored as a table of horizontal and vertical cut-off frequencies. The real-time processing module uses these values to design the optimal filter for the input image. The system expects that the content is stored in image-plus-depth format. The disparity value is used to select the corresponding column in each passband table. The user can set the value of the desired distortion level. This parameter is called "3D-sharpness" since it controls the trade-off between visibilities of details versus visibility of Moiré artefacts. The value of "3D-sharpness" is used to select the corresponding row of each table. The values in the selected cells give the desired vertical and horizontal cut-off frequencies of an anti-aliasing filter. These cut-off frequencies are used for designing the filters. More details of the framework can be found in [P06].

4.3 Content optimization

4.3.1 Crosstalk mitigation

In [74], Konrad *et al.* propose a pre-compensation algorithm for reducing the crosstalk in stereoscopic displays. However, their approach is not suitable for multiview displays; for these, pre-compensation mitigates the effect for a certain observation angle only, while amplifying it for other angles. As a multiview display is intended for many observers, it is desirable to mitigate the ghosting artefacts for all observation angles simultaneously. The straightforward approach to mitigate the crosstalk is to smooth the scene observations in the horizontal direction, where the level of smoothing depends on the amount of the parallax (i.e. disparity) [2]. For a scene in image-plus-depth format this corresponds to smoothing of the 2D image, with level of smoothing dependent on the absolute depth values of the pixels.

Two algorithms for crosstalk mitigation for multiview displays, which can be implemented in GPU, are discussed in [P10]. The first algorithm employs pre-filtering of the 2D image before using it as a texture on the mesh. It uses eight filters for the whole range of depth values. Eight

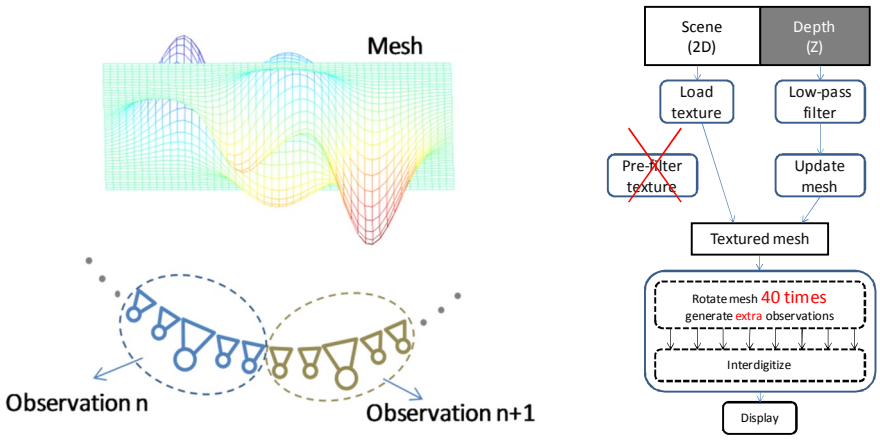


Figure 45, Crosstalk mitigation with pre-filtering: a) position of the extra observation points in respect to the original ones and b) block diagram of the algorithm. Appears in [P10], © 2008 SPIE, reprinted with permission.

masks are prepared, passing different ranges of depth values, according to the distance from the screen level. Each mask is applied to the corresponding filtered image and the result is blended together in the accumulation buffer. The algorithm is implemented using the CUDA library [85].

The second algorithm uses an image scattering technique for crosstalk mitigation. It works by blending extra observations with the ones needed for the multiview display. Around each observation point used in the previous approach, observation points are added at equal angles. The observation points are grouped as shown in Figure 45a. The images rendered from a group of observation points are blended together in a single image which is mapped to the sub-pixels belonging to one view of the screen. The algorithm works in a similar way to the previous one, with the exception that instead of pre-filtering the texture, additional observations are rendered as illustrated in Figure 45b. Additional information and visual examples can be found in [P10], included in this compound thesis.

4.3.2 Re-purposing

Excessive disparity is a problem most often found in 3D content which is created for one display size and is observed in another. Adapting the size and disparity of 3D content to fit a given 3D display is known as *content repurposing*.

In [86] an algorithm for content repurposing on a mobile device is discussed. An important

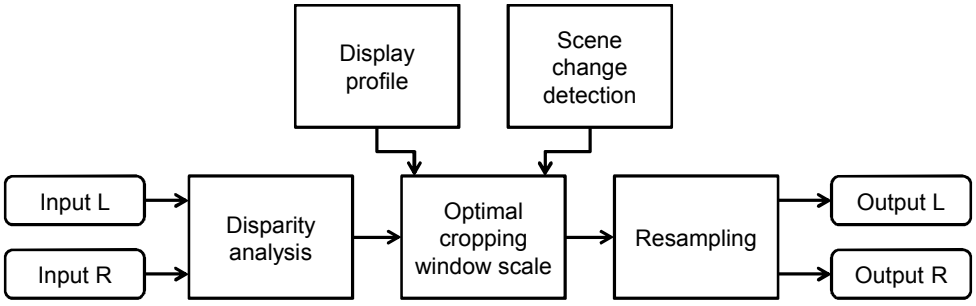


Figure 46, Block diagram of an algorithm for fast 3D content re-purposing.

requirement for such an algorithm is that it can be used for real time repurposing. Unfortunately, commonly used repurposing algorithms such as virtual view generation of non-liner disparity correction [75] are too computationally expensive to be used for real-time conversion on a contemporary portable device. In order to simplify the computation the algorithm uses *horizontal image translation* (H.I.T) which involves finding the size and position of a scaling window. In the H.I.T.-based repurposing algorithm one first finds the disparity of the source video, then finds the optimal cropping and scaling parameters, and then performs the actual image resampling. Having a dense disparity map is not critically important for performing H.I.T. – it is sufficient to know parameters of the disparity distribution, such as minimum, maximum and mean disparity.

The algorithm consists of five stages as shown in Figure 46. First, the comfort disparity range of the display is derived. Then the disparity range of the input content is calculated. Based on the estimated input and desired output disparity ranges, the algorithm derives the optimal scale of the cropping window, which would then yield the targeted disparity range and minimize the area of cropped and letterboxed content. Once the rescaling and cropping parameters are known, the framework performs resampling procedure with a desired, perceptually optimal performance in the frequency domain. More details on the algorithm performance can be found in [86].

5 Conclusion

Stereoscopic displays are intended to recreate a scene in three dimensions. Due to technical limitations some visual features of the scene are lost. The differences are interpreted by the HVS as artefacts. The missing information cannot be fully reconstructed, however due to the absence of a visual reference it is possible to make the distortions less visible. This dissertation discusses signal processing techniques to decrease the visibility of artefacts on a 3D display.

Visual optimization techniques require knowledge of both human vision and display design. One needs to know the important visual properties of a 3D scene and the relevant display properties that allow the scene to be shown in 3D. This dissertation proposes methods for 1) deriving visual properties of 3D displays, 2) predicting the visibility of artefacts and 3) visual optimization of 3D content. Visual quality estimation methods cover large glasses-enabled displays, small autostereoscopic displays, and dual-view and multiview displays. The proposed methods for visual optimization over distortions related to 1) observation position, 2) head pose, 3) view multiplexing and 4) excessive disparity range.

The dissertation is a compound of summary and original publications. In the summary, the problem was introduced, properties of the HVS and 3D displays were discussed and the scientific approach toward visual optimization of 3D content was presented along with the current state of the art. The second part of dissertation contains 4 journal and 8 conference publications on 3D artefact classifications, visual quality evaluation and various artefact mitigation techniques.

5.1 Results

The scientific results are presented in the publications attached to this thesis. A classification of artefacts in 3D content is proposed in [P02] and [P08]. The classification is based on the creation of distortions from various stages in 3D content transition and the interpretation of these distortions by processing subsystems in the HVS. A generalized model of a multiview display is proposed in [P01], [P04] and [P06]. The model is used to explain the reason for various 3D artefacts occurring. Visibility of artefacts is studied in [P01] (for multiview displays) and [P05] (for mobile 3D displays). A method for measuring the optical parameters of a multiview display is introduced in [P01], [P04] and [P11]. A method for deriving the so-called display passband is proposed in [P01] and [P06]. The passband allows easy comparison of the visual quality of multiview displays and also allows content producers to optimize content for a given display (or, select display optimal for a given content). In [P02], [P03], [P06] and [P07] various methods for building pass-band optimizing filters are proposed. The filters address visibility of Moiré, masking and some ghosting artefacts. In [P06] a tuneable system which allows the observer to select his or her preferred level of filtering is given. Results of optical measurements for mobile displays are presented in [P02] and [P05]. Finally, various algorithms for real-time visual optimization are proposed in [P07] (antialiasing for multiview displays), [P09] (extending head parallax for multiview displays), [P10] (GPU based crosstalk mitigation for multiview displays) and [P12] (viewpoint optimization for mobile displays). The algorithms described in [P09] and [P12] use real-time observer-tracking, which is implemented as a combination of face and eye-tracking, and is described in [P12].

5.2 Future work

Future work includes deeper study of temporal HVS properties and their effect on the perceived quality. Even though temporal masking is studied for 2D vision, its effects on binocular perception are not well known. In another study, the author and colleagues initiated research on temporal properties of binocular gaze [87] but it is still in a very early stage. As a result, temporal distortions and the corresponding artefacts in temporally-multiplexed 3D displays with active glasses are not studied. Effects on prolonged exposure of the HVS to temporally-interleaved are unknown. A stereoscopic image which flickers in anti-phase is potentially overloading the LGN and could have unpredictable results over time. Another interesting aspect of the quality that needs further study is the link between global motion and frame rate in stereoscopic setting (either spatially or temporally interleaved).

The research presented in this thesis covers 3 different models of multiview displays, two models of 3D displays with passive glasses, and 11 models of autostereoscopic displays. The results of passband optimization need to be validated using additional models of multiview displays. Passband optimization works best if different filters are applied for content with different disparity levels. The algorithms, presented in this thesis rely on the availability of a dense disparity map (or, assume average passband for all disparity levels). Deriving dense disparity map is computationally intensive. One alternative approach that needs to be studied is the use of 3-D filters over the epipolar volume representation of a 3D scene.

Finally, while this thesis was focused on measuring display parameters and mitigation of display-specific artefacts, a general study on quality of stereoscopic content is missing. There is an obvious need for general-purpose 3D quality metric which can be applied to all stages in 3D content transmission, instead of being used for studying the quality of 3D displays only.

Bibliography

[1] M. Levoy and P. Hanrahan, "Light Field Rendering," *Proc. ACM SIGGRAPH,* pp. 31-42, 1996.

[2] M. W. Halle, "Holographic stereograms as discrete imaging systems," in *Practical Holography VIII*, San Jose, CA.

[3] B. A. Wandel, Foundations of vision, Sunderland, Massachusetts, USA: Sinauer Associates, Inc, 1995.

[4] I. P. Howard and B. J. Rogers, Binocular Vision and Stereopsis, New York: Oxford University Press, 1995.

[5] D. Chandler, "Visual Perception (Introductory Notes for Media Theory Students," MSC portal site, University of Wales, Aberystwyth, 2008. [Online]. Available: http://www.aber.ac.uk/media/sections/image.html.

[6] J. Dowling, The Retina: an Approachable Part of the Brain, Harvard University Press, 1987.

[7] J. Ferwerda, "Elements of early vision for computer graphics," *IEEE Computer Graphics and Applications,* vol. 21, no. 5, pp. 22-33, Jul-Aug 2001.

[8] S. Winkler, Digital Video Quality, John Wiley & Sons, 2005.

[9] Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing,* vol. 13, no. 4, pp. 600-612, 2004.

[10] S. Lee, M. S. Pattichis and B. A. C., "Foveated video compression with optimal rate control," *IEEE Transactions on Image Processing,* vol. 10, pp. 977-992, July 2001.

[11] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color image reproduction," *Journal of the SID5,* vol. 5, no. 1, pp. 61-63, 1997.

[12] ISO/IEC JTC 1/SC 29/WG 1, "Coding of Still Pictures," [Online]. Available: http://www.itscj.ipsj.or.jp/sc29/29w12901.htm.

[13] K. M. Schreiber, J. M. Hillis, H. R. Filippini, C. M. Schor and M. S. Banks, "The surface of the empirical horopter.," *Journal of Vision,* vol. 8, no. 3, pp. 1-20, 2008.

[14] E. D. Montag and M. D. Fairchild, "Fundamentals of Human Vision and Vision Modelling," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. H. Rao, Eds., Boca Raton, FL, CRC Press, 2006, pp. 45-81.

[15] S. Pastoor, "Human factors of 3D imaging: Results of recent research at Heinrich- Hertz-Institut Berlin," in *2nd International Display Workshop*, Hamamatsu, 1995.

[16] D. B. Diner, "A new definition of Orthostereopsis for 3-D Television," in *IEEE International Conference on Systems, Man and Cybernetics*, 1991.

[17] M. Wexler and J. Boxtel, "Depth perception by the active observer," *Trends in Cognitive Sciences,* no. 9, pp. 431-438, 2005.

[18] B. Julesz, Foundations of Cyclopean Perception, Chicago: The University of Chicago Press, 1971.

[19] E. Stoykova, A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar and X. Zabulis, "3-D Time-Varying Scene Capture Technologies—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 17, no. 11, pp. 1568-1586, 2007.

[20] P.-S. Tsai, J. Cryer and M. Shah, "Shape-from-shading: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 21, no. 8, pp. 690 - 706 , 1999.

[21] T. Lindeberg and J. Garding, "Shape from texture from a multi-scale," in *ICCV*, 1993.

[22] M. Subbarao and G. Surya, "Depth from Defocus: A Spatial Domain Approach," *International Journal of Computer Vision,* vol. 13, pp. 271-294, 1994.

[23] A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, 2004.

[24] M.-H. Yang, D. Kriegman and N. Ahuja, "Detecting faces in images: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 24, no. 1, pp. 34-58, 2002.

[25] H. Sidenbladh, M. Black and L. Sigal, "Implicit probabilistic models of Human Motion for Synthesis and Tracking," in *European Conference on Computer Vision*, 2002.

[26] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in *Proc. Comput. Vis. and Pattern Recognit. (CVPR2006)*, 2006.

[27] S. e. al., "Research progess on scannerless ladar systems using a laser diode transmitter and FM/cw radar principles," in *Laser Radar Technology and Applications VI*, 2001.

[28] U. Schnars and J. W., "Direct recording of holograms by a CCD target and numerical reconstructions," *Applied Optics,* vol. 33, no. 2, pp. 179-181, 1994.

[29] A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, E. C. and A. Weigel, "Scene Representation Technologies for 3DTV—A Survey," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 17, no. 11, pp. 1587-1605, Nov. 2007.

[30] M. Halle, "Multiple Viewpoint Rendering," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998.

[31] R. Hartly and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. ed., Cambridge University Press, 2006.

[32] A. Smolic, K. Mueller, N. Stefanovski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis and A. Koz, "Coding Algorithms for 3DTV - A Survey," *IEEE Trans. Circuits and Systems for Video Technology,* vol. 17, no. 11, pp. 1606-1621, Nov. 2007.

[33] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek and I. Sexton, "An evolutionary and optimized approach on 3D-TV," in *Int. Broadcast Conf.*, Amsterdam, The Netherlands, 2002.

[34] C. Fehn, "3D-TV using depth-image-based rendering (DIBR)," in *Picture Coding Symp.*, San Francisco, CA, USA, 2004.

[35] C. Fehn, N. Atzpadin, M. Muller, O. Schreer, A. Smolic, R. Tanger and P. Kauff, "An Advanced 3DTV Concept Providing Interoperability and Scalability for a Wide Range of Multi-Baseline Geometries," in *2006 IEEE International Conference on Image Processing*, 2006.

[36] R. Fernando and M. J. Kilgars, The Cg Tutorial, The Definitive Guide to Programmable Real-Time Graphics, Addison-Wesley, 2006.

[37] J. Lee, "Hacking the Nintendo Wii Remote," *Pervasive Computing, IEEE,* vol. 7, no. 3, pp. 39-45, 2008.

[38] K. Akeley, S. J. Watt, A. R. Girshick and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.,* vol. 23, no. 3, p. 804–813, 2004.

[39] M. Saymta, S. Isikman and H. Urey, "Scanning Led Array Based Volumetric Display," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008.

[40] S. Pastoor, "3D displays," in *3D video Communication*, O. Scheer, P. Kauff and T. Sikora, Eds., Chichester, West Sussex, Wiley, 2005, pp. 235-260.

[41] P. Surman, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanar and J. Watson, "Solving the 3D problem - The history and development of viable domestic 3-dimensional video displays," in *Three-Dimensional Television: Capture, Transmission, and Display*, H. Ozaktas and L. Onural, Eds., Springer Verlag, 2007.

[42] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov and C. von Kopylow, "A Survey of 3DTV Displays: Techniques and Technologies," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 17, no. 11, pp. 1647-1658, Nov. 2007.

[43] H. Urey, K. V. Chellappan, E. Erden and P. Surman, "State of the Art in Stereoscopic and Autostereoscopic Displays," *Proceedings of the IEEE,* vol. 99, no. 4, pp. 540-555., 2011.

[44] L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanar and J. Watson, "An Assessment of 3DTV Technologies," in *NAB Broadcast Engineering Conference Proceedings*, Las Vegas, USA, 2006.

[45] H. Jorke, H. Simon and M. Fritz, "Advanced stereo projection using interference filters," *J. Soc. Inf. Display, ,* vol. 17, no. 5, pp. 407-410, 2009.

[46] Toshiba Europe GmbH, "55ZL2 - 3D wihout glasses," Toshiba, Jan 2012. [Online]. Available: http://eu.consumer.toshiba.eu/en/products/tv/55ZL2. [Accessed June 2012].

[47] W. L. IJzerman, S. T. de Zwart and T. Dekker, "Design of 2D/3D switchable displays," *Proc.*

of the SID, vol. 36, no. 1, pp. 98-101, May 2005.

[48] C. van Berkel and J. Clarke, "Characterisation and optimisation of 3D-LCD module design," in *Stereoscopic Displays and Virtual Reality Systems IV*, San Jose, 1997.

[49] W. Tzschoppe, T. Brueggert, M. Klipstein, I. Relke and U. Hofmann, "Arrangement for two-or-three-dimensional display". US Patent 2006/0192908, 31 Aug. 2006.

[50] N. Dodgson, "Autostereoscopic 3D Displays," *Computer,* vol. 38, no. 8, pp. 31-36, Aug. 2005.

[51] C. van Berkel, "Lenticular screen adaptor". US Patent 6801243, 5 Oct. 2004.

[52] J. Konrad and P. Angiel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Processing,* vol. 15, no. 1, pp. 128-140, 2006.

[53] M. Zwicker, W. Matusik, F. Durand, H. Pfister and C. Forlines, "Antialiasing for automultiscopic 3D displays," in *ACM SIGGRAPH 2006*, Boston, Massachusetts, 2006.

[54] V. Saveljev, J.-Y. Son, B. Javidi, S.-K. Kim and D.-S. Kim, "Moiré minimization condition in three-dimensional image displays," *Display Technology,* vol. 1, pp. 347-353, 2005.

[55] C. N. Moller and A. R. L. Travis, "Correcting interperspective aliasing in autostereoscopic displays," *, IEEE Trans. Visual Comput. Graphics,* vol. 11, no. 2, pp. 228-236, 2005.

[56] Merriam-Webster, Webster's Encyclopedic Unabridged Dictionary of the English Language, Gramercy , 22 April.

[57] M. Halle, "Autostereoscopic displays and computer graphics," in *International Conference on Computer Graphics and Interactive Techniques*, 2005.

[58] M. Yuen, "Coding Artefacts and Visual Distortions," in *Digital Video Image Quality and Perceptual Coding*, H. Wu and K. Rao, Eds., CRC Press, 2005.

[59] "A survey of MC/DPCM/DCT video coding istortions," *Signal Processing,* vol. 70, no. 3, p. 247–278, Nov. 1998.

[60] W. IJsselsteijn, P. Seuntiens and L. Meesters, "Human factors of 3D displays," in *3D Video Communication*, Scheer, Kauff and Sikora, Eds., Wiley, 2005, pp. 219-233.

[61] F. Kooi and A. Toet, "Visual comfort of binocular and 3D displays," *Displays,* vol. 25, no. 2-3, pp. 99-108, 2004.

[62] D. Hoffman, A. Girshick, K. Akeley and M. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision,* vol. 8, no. 3, pp. 1-30, 2008.

[63] A. Jain and J. Konrad, "Crosstalk on automultiscopic 3-D displays: Blessing in disguise?," in *Stereoscopic Displays and Applications XVIII, IS&T/SPIE Electronig Imaging*, San Jose, CA, 2007.

[64] M. Salmimaa and T. Jarvenpaa, "Optical characterization of autostereoscopic 3-D displays," *J. Soc. Inf. Display,* vol. 16, no. 825, 2008.

[65] J. Hakkinen, J. Takatalo, M. Kilpelainen, M. Salmimaa and G. Nyman, "Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width," *J. Soc. Inf. Display,* vol. 17, no. 433, 2009.

[66] S. K. Nayar, V. Branzoi and T. E. Boult, "Programmable Imaging Using a Digital Micromirror Array," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2004.

[67] B. T. Bakus, M. S. Banks, R. van Ee and J. A. Crowell, "Horizontal and vertical disparity, eye position, and stereoscopic slant perception," *Vision Research,* vol. 39, pp. 1143-1170, 1999.

[68] P. Boher, T. Leroux, T. Bignon and V. Collomb-Patton, "A new way to characterize auto-stereoscopic 3D displays using Fourier optics instrument," in *Stereoscopic Displays and Applications XX, SPIE 7237, 72370Z*, 2009.

[69] S. Uehara, T. Hiroya, H. Kusanagi, K. Shigemura and H. Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement," in *Stereoscopic displays and applications XIX*, 2008.

[70] P. Debevec and M. J., "Recovering High Dynamic Range Radiance Maps from Photographs," in *ACM Siggraph*, 1997.

[71] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision," in *SPIE Photonics West 2002: Electronic Imaging*, 2002.

[72] M. Salmimaa and T. JArvenpaa, "3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays," *Soc. Inf. Display,* vol. 16, p. 1033, 2008.

[73] P. Surman, R. Brar, I. Sexton and K. Hopf, "MUTED and HELIUM3D autostereoscopic displays," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, July 2010.

[74] J. Konrad, B. Lacotte and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Process.,* vol. 9, pp. 897-908, May 2000.

[75] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic and M. Gross, "Nonlinear Disparity Mapping for Stereoscopic 3D," *ACM Transactions on Graphics (Proc. SIGGRAPH),* vol. (in press), 2010.

[76] Heinrich Hertz Institute, "Free2C Desktop Display," 2012. [Online]. Available: http://www.hhi.fraunhofer.de/en/departments/interactive-media-human-factors/overview/free2c-desktop-display/. [Accessed 10 June 2012].

[77] P. Surman, I. Sexton, K. Hopf, W. K. Lee, F. Neumann, E. Buckley, G. Jones, A. Corbett, R. Bates and S. Talukdar, "Laser-based multi-user 3-d display," *J. Soc. Inf. Display,* vol. 16, pp. 743-753, 2008.

[78] R. Brar, P. Surman, I. Sexton, R. Bates, W. Lee, K. Hopf, F. Neumann, S. Day and E. Willman, "Laser-Based Head-Tracked 3D Display Research," *Display Technology, Journal of,* vol. 6,

no. 10, pp. 531-543, 2010.

[79] K. Hopf, F. Neumann and D. Przewozny, "Multi-user eye tracking suitable for 3D display applications," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011*, 2011.

[80] V. Uzunov, A. Gotchev, K. Egiazarian and J. Astola, "Face Detection by Optimal Atomic Decomposition," in *SPIE Optics and Photonics 2005: Algorithms, Architectures, and Devices and Mathematical Methods, Mathematical Methods in Pattern and Image Analysis*, San Diego, California, USA, 2005.

[81] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis,* vol. 10, no. 3, pp. 234-253, May 2001.

[82] H. Essaky Sankaran, A. Gotchev, K. Egiazarian and J. Astola, "Complex wavelets versus Gabor wavelets for facial feature extraction: a comparative study," in *Proc. SPIE Image processing : algorithms and systems IV, Vol. 5672*, San Jose, CA, 2005.

[83] A. Boev, M. Georgiev, A. Gotchev and K. Egiazarian, "Optimized single-viewer mode of multiview autostereoscopic display," in *Proc. of 16th European Signal Conference EUSIPCO 2008*, Lausanne, Switzerland, 2008.

[84] S. K. Mitra, Digital signal processing: A computer based approach, 3 ed., New York: McGraw-Hill, 2005.

[85] V. Podlozhnyuk, "Image Convolution with CUDA, white paper," Nvidia Corp, June 2007. [Online]. Available: http://developer.download.nvidia.com/compute/cuda/sdk/website/projects/convolutionSeparable/doc/convolutionSeparable.pdf. [Accessed June 2012].

[86] A. Karaoglu, B. H. C. W.-S. Lee, A. Boev and A. Gotchev, "Fast repurposing of high-resolution stereo video content for mobile use," in *Real-Time Image and Video Processing 2012,* , Brussels, Belgium, 2012.

[87] A. Boev, M. Hanhela, A. Gotchev, T. Utriainen and S. Jumisko-Pyykko, "Parameters of the human 3D gaze while observing portable autostereoscopic display: a model and measurement results," in *Multimedia on Mobile Devices 2012*, San Francisco, CA, USA, 2012.

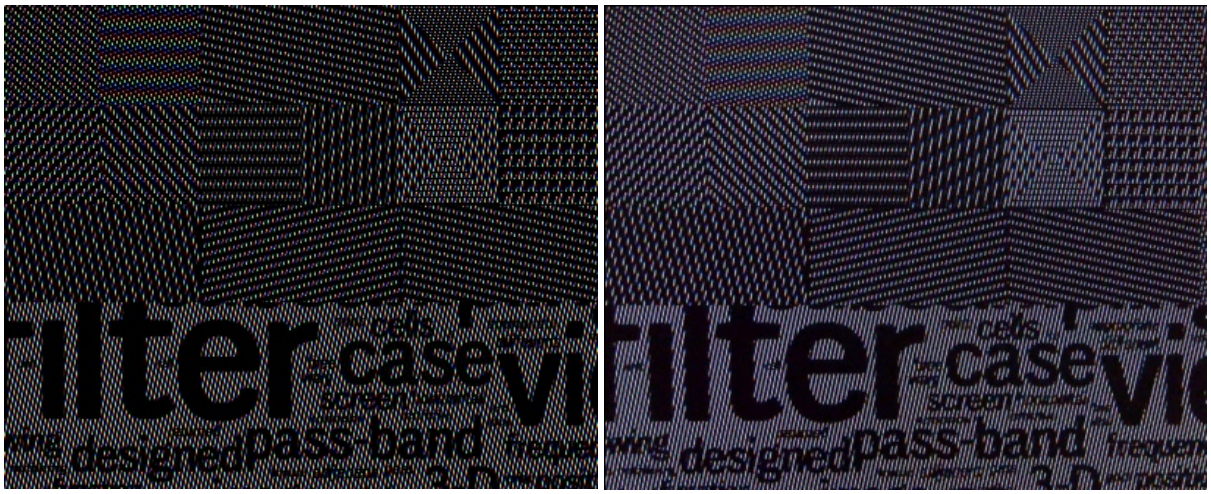
[88] "Kodak Image Database," Kodak, [Online]. Available: ftp://ftp.kodak.com/www/images/pcd .

Appendix I: Test content, visualised on 3D displays

The purpose of this appendix is to provide detailed, full-colour reproduction of some images, which are meant for visual comparison. Most of these are test images, visualized and photographed on various 3D displays.

In Figure 47 one can see an example of simulated multiview display output. The original test image is shown in Figure 47a. Simulated display output of the same test image is shown in Figure 47b. The simulation is based on the measurement methodology for finding the interleaving pattern and deriving the angular brightness function of a sub-pixel proposed in [P04]. The actual display output obtained by photographing the display is given in Figure 47c. More information can be found in Section 3.1.

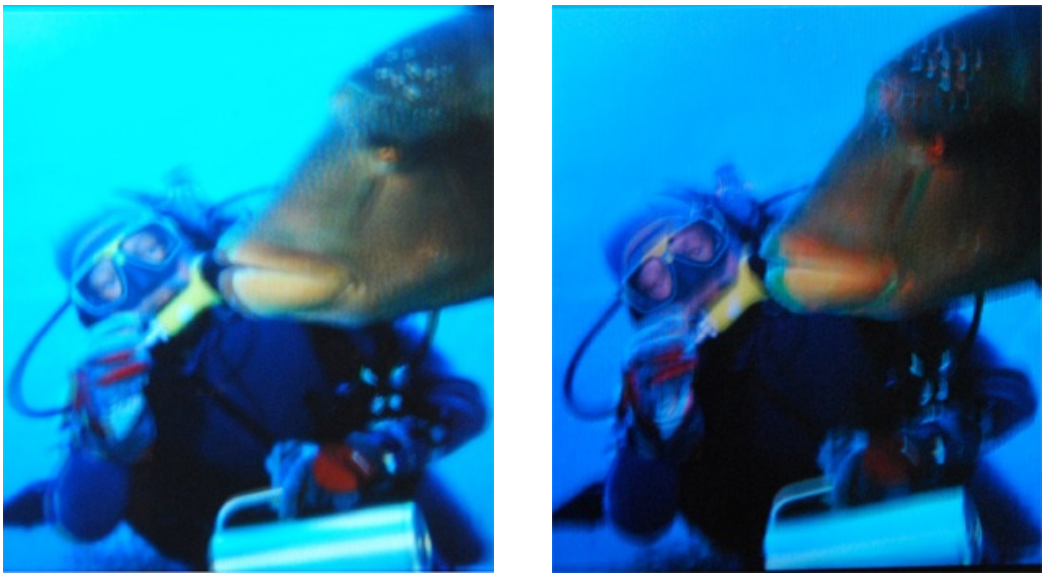
a)



b) c)

Figure 47. Simulation of display output: a) origintal test image, b) simulated screen output, c) actual screen output, photographed

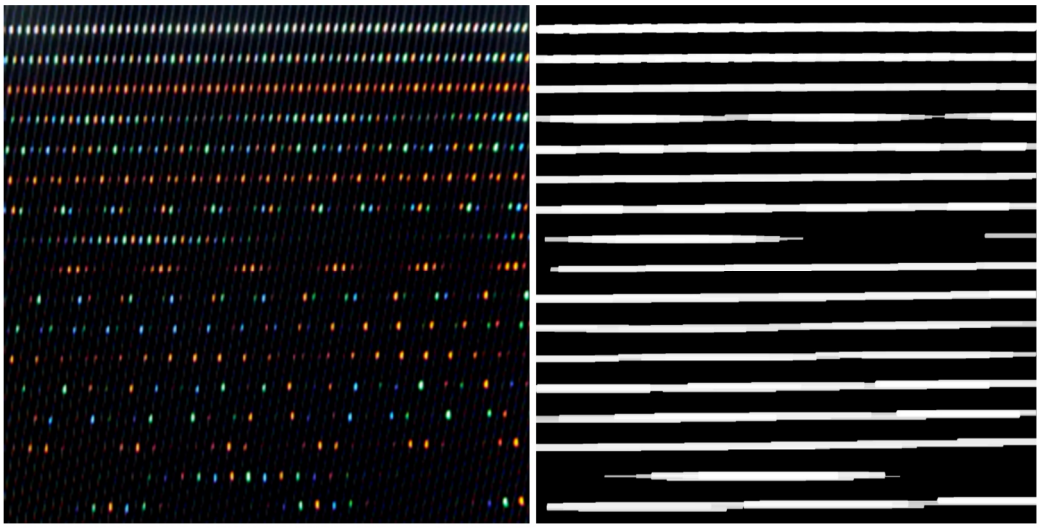
An example of ghosting artefacts is shown in Figure 48. Both images in the figure are prepared by photographing a 3D display with a camera placed in its stereo-edge. Figure 48a is a photograph of an autostereoscopic display with sub-pixel interleaving and gives an example for colour-balanced ghosting artefacts. Figure 48a is a photograph of an autostereoscopic display with pixel interleaving, and gives an example for ghosting artefacts, combined with colour-bleeding. More information can be found in Section 3.1.



a) b)

Figure 48. Ghosting artefacts and colour bleeding: a) ghosting artefacts with no colour bleeding, b) ghosting artefacts with colour bleeding

Example close shot for pixel groups with various N/V ratios is shown in Figure 49. Each line in Figure 49a has its every *n*-th sub-pixel turned on, where *n* varies from 1 to 17. In Figure 49b one can see the N/V regions, obtained by filtering each line with median filter with length *n*. More information can be found in Section 3.2.



a) b)

Figure 49. Close shot of pixel groups with various N/V ratio: a) lines with very *n*-th sub-pixel turned on, for *n*=1..17, b) N/V regions for this lines.

An example of anti-aliasing filters for natural images can be seen in Figure 50. In Figure 50a one can see the original test image (Lighthouse, taken from Kodak Image Database [88]). Photographs of a multiview display showing the same test image are given in Figure 50b-d. Figure 50b shows the image without pre-processing. Figure 50c shows an image, pre-filtered with subsampling-pattern-aware filter, based on the methodology proposed in [52]. Figure 50d shows the image pre-filtered with a 2D passband filter, as described in Section 4.2.1.

a) b)

c) d)

Figure 50. Antialiasing filtering of natural images for autostereoscopic displays: a) original test image, b)-d) photographs of a multiview display, b) unprocessed image, c) pre-filtered with subsampling-pattern-aware filter (porposed in [52]), d) pre-filtered with a 2D passband filter (proposed in [P07])

An example of anti-aliasing filters optimized for text is given in Figure 51. In Figure 51a one can see the original test image Photographs of a multiview display showing the same test image are given in Figure 51b-d. Figure 51b shows the image without pre-processing. Figure 51c shows an image, pre-filtered with subsampling-pattern-aware filter, based on the methodology proposed in [52]. Figure 51d shows the image pre-filtered with a 2D passband filter, as described in Section 4.2.1. Figure 51e shows the image pre-filtered with a non-separable filter with non-square passband. Figure 51f shows the image pre-filtered with a non-separable filter with non-square passband. More information can be found in Section 4.2.1 and Section 4.2.2.

a) b)

c) d)

e) f)

Figure 51. Antialiasing filtering of text for autostereoscopic displays: a) original test image, b)-f) photographs of a multiview display, b) unprocessed image, c) pre-filtered with subsampling-pattern-aware filter (porposed in [52]), d) pre-filtered with a 2D passband filter (proposed in [P07]), e) pre-filtered with separable filter with a non-square passband (discussed in [P03]), f) pre-filtered with separable filter with a square passband (proposed in [P03]).

Appendix II: Author's contribution to the publications

The contribution of the author per publication is the following:

P01 – Sections 1, 2, 3, 5 and 6 were written by A. Boev.

P02 –A. Boev contributed to Section II.A, major part of Sections IIIa and IIIb, major part of Section IV.A, and parts of Section V.

P03 – Sections 1, 2 (excluding subsection 2.5), 4 and 5 were written by A. Boev.

P04 – Sections 1, 2, 3, 4, and 6 were written by A. Boev.

P05 – The paper was written by A. Boev.

P06 – Sections 1, 2 (excluding subsection 2.3) 3, 4., 7 and 8 were written by A. Boev.

P07 – Sections 1, 2 and 5 were written by A. Boev.

P08 – Sections 1, 2, 3, 4, part of Section 5, and 7 were written by A. Boev.

P09 – Sections 1, 2, 3, 6, and 7 were written by A. Boev.

P10 – The paper was written by A. Boev.

P11 – The paper was written by A. Boev.

P12 – Sections 2 and 3 were written by A. Boev.

Original publications

[P01] A. Boev, R. Bregovic and A. Gotchev, "Visual-quality evaluation methodology for multiview displays," *Displays,* vol. 33, no. 2, pp. 103-112, April 2012.

This page left blank

Visual quality evaluation methodology for multiview displays

Atanas Boev, Robert Bregovic, Atanas Gotchev

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

firstname.lastname@tut.fi

ABSTRACT

Multiview displays are characterized by a multitude of parameters such as spatial resolution, brightness, 3D-crosstalk, etc., which individually and in their combination influence the visual quality of the displayed 3D scene. These parameters are specified by values, precisely measured by optical means. However, it is difficult for an average consumer or content producer to compare the visual quality of two displays or judge if a given 3D content is suitable for a certain display only by this set of parameter values. In this paper, we propose a quality measurement methodology, which aims at measuring the visibility of structural distortions, introduced by a multiview display, to a number of test signals with different frequency content and apparent depth. We use these measurements to derive what we call *display passband* for signals at different disparity levels. The passband determines the frequency components for which the intended signal is predominantly visible, with respect to the distortion introduced by the display. Additionally, we propose a method to determine the approximate effective resolution of the display for signals with a given apparent depth. The result of the measurements can be used to compare the perceived visual quality of different multiview displays.

1. INTRODUCTION

Multiview displays can create visual illusion of objects floating in 3D space without requiring the observer to wear 3D glasses. Typically, multiview displays combine a pixel-addressable matrix, such as in plasma, LED, or LCD panels, with additional *optical layer* mounted on top [1]. The optical layer redirects the light generated by the pixel matrix, making the visibility of each pixel element a function of the observation angle. The set of elements visible from certain angle forms an image, also called a *view* [1][2]. A multiview display can simultaneously show a number of different views, each one visible from different direction. The process of combining multiple images in one compound bitmap is referred to as *interleaving*, and the map that links the position of TFT elements with the view number they belong to is referred to as *interleaving map*. If the views are properly selected observations of the same scene, the display recreates the scene in 3D. Even though all objects of the scene are projected on the display, they might appear as they are at different distances to the observer. The apparent distance to an object is referred to as its *apparent depth*. If the object

appears at the display level, all its observations appear on the same display coordinates. If the object has different apparent depth, it appears on different horizontal coordinates in each view. The distance between the positions of an object in different views is referred to as *disparity*. The objects with positive disparity appear behind the display level, and those with negative disparity appear in front of the display.

The downside of the optical layer is that it introduces a number of multiview display specific artifacts [1]. In addition to monoscopic display parameters, such as 2D resolution or refresh rate, the visual quality of a 3D monitor is influenced by variables such as 3D-contrast and 3D-crosstalk [3]. The multitude of parameters hinders the comparison of the visual quality of different multiview displays. A number of previous works have addressed the estimation of display optical quality, ranging from theoretical considerations about the interleaving map [4][5][6] trough measuring of the optical parameters of the display [3][7][8] to subjective tests with different multiview displays [9][10][11]. However, evaluating the quality of a multiview display based on its optical parameters only, has two main disadvantages – first, some parameters, e.g. luminance uniformity across different observation angles, are not directly related to the perceived quality; and second, visibility of 3D artifacts depends also on scene content, observation conditions and properties of the human visual system. Having human observers to rate the visibility of artifacts in all scenes would be an optimal quality assessment approach; however it is expensive and time-consuming.

In this article we propose a methodology aimed at evaluating the level of signal distortion introduced by a given multiview display. We use multiple test signals with various frequency components and disparity levels to derive the display passband regions for planes with different apparent depth. The passband regions are estimated for two levels of distortion visibility. The shapes of the passband regions can be used to estimate the expected visual quality for different types of 3D content. Additionally, we devise a method aimed at approximating the equivalent resolution of the display for given disparity range and distortion level. The equivalent resolution of a multiview display can be directly used as a perceptually-relevant indicator of its visual quality. In a previous work, we have proposed passband evaluation methodology, which did not consider disparity [12]. Here, we extend our previous approach for a range of disparities and multiple distortion levels. In another work [13], we have estimated the distortion levels based on knowledge of interleaving pattern and angular visibility function. However, we noticed that non-linear optical effects are introducing significant differences in passband regions in measured versus simulated data. The results in the current article are based solely on measurements, and do not require knowledge of the angular visibility.

The paper is structured as follows: in Section 2, we introduce the concept of distortion visibility to be used as indicator of perceptual quality. We introduce a model of multiview display, use it to explain the most common artifacts, and give a general methodology for measuring and evaluating visual distortions. In Section 3, we give details about the measurement procedure, like test image preparation and experimental setup. We give a practical example with measurements of a 24-view display. In Section 4, we explain how the measured data is evaluated in order to obtain the

passband regions of the display for a given distortion level. In Section 5 we show how using these regions one can approximate the equivalent resolution of the display for different depth planes and different distortion levels. In the last section we give concluding remarks.

2. VISIBILITY OF DISTORTIONS AS INDICATOR OF VISUAL QUALITY

2.1 Artifacts in 3D displays

The most pronounced artifacts in a multiview display are moiré and ghosting artifacts [16]. Typically, the visible pixels of a view appear on a non-orthogonal grid [1][4]. Mapping the input images, which are usually sampled on rectangular grid, to the visible pixels of a view requires special anti-aliasing filters [5][6][17]. Direct mapping of multiple images to the views of a multiview display produces moiré and color aliasing artifacts similar to the ones shown in Figure 1a. The design of a multiview display involves a trade-off between number of views, spatial resolution of a view, and visibility artifacts such as *image flipping* and *banding* [4][11]. Often, the visibility zones of different views are interspersed and from a given angle multiple views are simultaneously visible, albeit with different brightness [1][2][3]. When visualizing 3D objects with pronounced depth the combination of disparity and simultaneous visibility is perceived as ghosting artifacts [5][10]. An example for ghosting artifacts is given in Figure 1b. Often, the process responsible for ghosting is modeled as crosstalk, and the term crosstalk is used as a synonym for ghosting artifacts [3][5][10][11][16]. For displays with parallax barrier, the barrier creates visible gaps between the pixels, as seen in Figure 1c. These gaps are seen as masking artifacts [12], similar to the fixed-pattern noise exhibited by some digital projectors [18].

a) **b)** **c)**

Figure 1 Typical artifacts exhibited by multiview displays: a) moiré, b) ghosting and c) masking

2.2 Multiview display as an image processing channel

In order to assess perceptual differences between intended (input) and visualized (output) signal, we propose a model, which considers the multiview display as an image processing channel. The model follows the steps of content

preparation and visualization, and has five stages as shown in Figure 2. For simplicity, the examples in the figure are given for single row of a dual-view display, however, the general signal transformations hold true for any multiview display. The channel input is an image, which is meant to be seen at a particular depth. It can be regarded as a continuous signal, as it is shown in Figure 2a. The second stage models the preparation of the views. From the interleaving map one can derive a binary mask defining the position of the samples in one view. The input signal is sampled using the binary mask of the first view. For example, let these samples appear of positions with odd number as horizontal coordinate, as shown in Figure 2b. The third stage models the presence of disparity. If the input image is meant to be seen on the level of the display, it appears on the same place in each view. In that case, the same input signal is sampled using the binary mask of the second view. For example, positions with even number as horizontal coordinate, as seen in Figure 2c. If the input image should appear at different depth, it has disparity between the views, and an offset version of the input signal is sampled, as exemplified in Figure 2f. The fourth stage models the process of interleaving. Following the interleaving map, observations of the same object with different disparity are combined together. In our model, this corresponds to a combination of multiple versions of the same input signal, sampled with different offset. For example, an interleaved version of the input signal with zero disparity is shown in Figure 2d. It is made by alternating the samples in Figure 2b (odd positions) and the ones in Figure 2c (even positions). Alternatively, an example for the same input signal interleaved with disparity 20 is given in Figure 2g, which is a combination of the samples in Figure 2b and Figure 2f. The last stage models the influence of the optical layer. The layer changes the visibility of the individual display elements depending on the observation angle. In our example, from certain observation position the odd samples are seen with full brightness, while the even samples are seen with one quarter of the brightness (Figure 2e and Figure 2h).

It should be noted, that for objects with zero disparity the interleaved image (Figure 2d) is a good representation of the input signal (Figure 2a). In that case, a multiview display can be modeled as a 2D display where parts of the image have partial visibility, as suggested by Jain and Konrad in [14]. The less the impact of the optical layer is, the closer the visual output to the input signal is (Figure 2e), and – as Jain and Konrad have proven – the bigger the frequency throughput of the display is. Lower visibility of the masked pixels results in alternating bright and dark pixels (Figure 2e), which can be modeled as fixed-pattern noise. However, if disparity is introduced, the interleaved signal (Figure 2g) is quite different than the input signal. In that case, the masking effect of the optical layer makes the output (Figure 2g) better representation of the input signal. The shifted version of the input signal is meant to be fully visible from another observation angle. If it is partially present in the current observation angle, as shown in Figure 2h, it is modeled either as crosstalk between the views [7], or as interperspective aliasing [15].

The sampling stage in Figure 2 imposes an anti-aliasing filter before it. We deliberately do not include it into the model. In a multiview display, beside aliasing one has to simultaneously deal with other sources of distortions, such as imaging

and crosstalk. In order to simplify the evaluation methodology we do not apply any anti-aliasing filter to the images displayed for measurements. Thus, we would see clearly the aliasing artifacts along with other artifacts as well as their interaction. From a pre-processing filter implementation point of view, this would allow addressing most of distortions by a single filter. In other words, we deal with distortions according to their visibility regardless of their origin. As seen from the model, the only place one can influence the signal is before the sampling stage as this is the only place where aliasing can be eliminated. Filters designed by the proposed methodology would generally act as anti-aliasing filters but also cancel some other frequency being source of imaging and cross-talk artifacts. Note that by measuring the artifacts in this way one can design more-restrictive or less-restrictive filters depending on the interaction between artifacts. Also the measurements can quantify in a better way the limits in changing the filter parameters for providing subjectively more pleasant visualization.

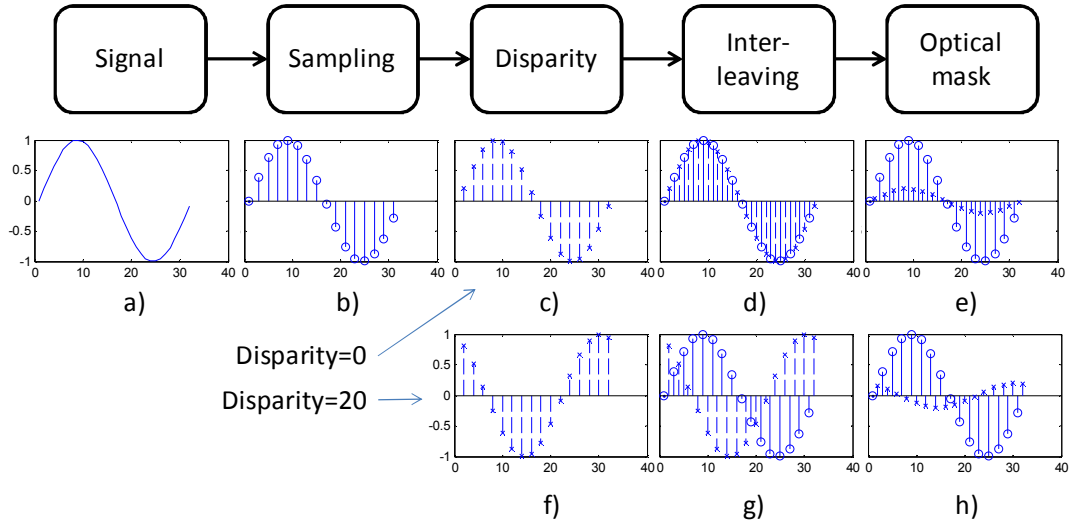


Figure 2 Generalized model of a multiview display as an image processing channel: a) input signal, b) input signal sampled in the positions which belong to view "1", c) input signal sampled in positions, which belong to view "2", d) interleaved signal, containing samples from views "1" and "2" and no disparity between the views, e) interleaved signal from "d)", after being masked by the optical layer, f) input signal, sampled in positions which belong to view "2", with an offset of 20 samples, g) interleaved signal, containing samples from views "1" and "2" and 20 samples disparity between the views, and f) interleaved signal from "g)", after being masked by the optical layer.

2.3 Properties of the human visual system

Most visual quality metrics work by assessing the perceptual difference between two images – one is the reference image and the other is the processed one. The reference is assumed to be of highest quality and the bigger the perceptual difference between the images is, the lower the quality of the processed image is deemed to be [20]. In the general case, however, the observer does not have the reference image available for comparison, and predicting the visibility of an artifact becomes a more complex task.

The act of seeing an object is a product of two phenomena – *visual perception*, which is the ability of the eye to collect visual data, and *visual cognition*, which is the ability of the brain to interpret visual information. The visual perception is limited by physiological factors, and these are modeled as *contrast sensitivity function* (CSF) [21]. The visibility of an image detail is influenced both by parameters of the vision and observation conditions [21]. In a stereoscopic image, the presence of crosstalk is additional factor which affects visibility of image details [10],[11]. According to the Weber-Fechner law the perceptibility of a change in stimuli is proportional to the amplitude of the stimuli. This fact also holds true for perception of brightness [19]. Following the Weber-Fechner law, the crosstalk is measured as percentage of the intended signal (intended signal is the input signal as perceived on the display). Crosstalk of less than 5% is considered under the visibility threshold and crosstalk of 25% or more is considered unacceptable [10]. The threshold level of barely acceptable crosstalk depends on the local contrast of the content and on the white-to-black contrast ratio of the display. This level has been reported as 10% [23], 15% [24] and 25% [10]. Typically, the level is measured with high-contrast, black-and-white patterns, but for natural images, a higher crosstalk level might be acceptable [23], [24]. In our paper, we use 20% as the value for barely acceptable crosstalk.

On the other hand, HVS is able to reconstruct the underlying structure even if it is partially obscured. The ability of the brain to recover occluded shapes and repetitive patterns is known as the *visual Gestalt principle* [19] , and the interdependent visibility of patterns with different properties is modeled as *pattern masking* [21]. According to the Gestalt principle, closely positioned shapes are grouped according to their proximity.

2.4 Criteria for visibility of distortions in multiview displays

In order to estimate the visibility of distortions, we model three HVS properties – brightness perception, contrast sensitivity function and Gestalt principle. This is done by a three-step procedure in frequency domain. First, we model the contrast sensitivity function by applying a circular weighting window. The weights in the window change as a function of the distance to the center of the coordinate system, and the shape of the function follows the shape of the spatial CSF at photopic level as described in [22]. Then, according to the Gestalt principle, we identify the visually dominant pattern by searching for the lowest spatial frequency regardless of the orientation. In frequency domain this is expressed as proximity of the peak of the signal to the center of the coordinate system (DC). Finally, according to the Weber-Fechner law, the eye senses brightness approximately logarithmically for typical observation conditions. Thus, we measure the visibility as the ratio between the amplitude of the distortion introduced by the display and the amplitude of the intended signal.

$$\delta(f_x, f_y) = \frac{\text{amplitude of the distortion}}{\text{amplitude of the signal}} \cdot 100 \ \%$$

In the rest of paper this is referred to as the *distortion to signal ratio*. Since the display behaves differently for different frequencies, the display distortion will depend on the horizontal, f_x, and vertical, f_y, signal frequency.

In this work we analyze the distortion by applying threshold at two different levels – 5% distortion level, which represents unnoticeable levels of distortion, and 20% which represents visible, but still acceptable artifact levels.

2.5 Evaluation methodology

In order to assess the visual quality of a multiview display, we prepared test images with varying spatial frequency, orientation and depth, and for each test image we measured the relative distortion introduced by the display. Our measurement methodology has six steps, which are shown in Figure 3.

The first step is to prepare number of *test signals*, which contain a 2D sinusoidal pattern with varying horizontal and vertical frequency components. Then, each test signal is extended to a number of *test images* each one with different apparent depth. This is done by mapping the same signal to each view of the display, adding different amount of disparity to each view and interleaving all views in a test image. The third step involves automated visualization of all test images on the display and making a snapshot of each one with a high resolution camera. The output of that step is a collection of *test shots* of all test images. In the next step the *spectra* of each test shots are analyzed, in order to determine the distortion to signal ratio, that is, the ratio between the magnitude of the distortion frequency component introduced by the display and the magnitude of the intended frequency component in the test signal. The distortion frequency component is selected as the largest peak in the spectra, which is positioned closer to the center than the intended frequency component of the input signal. Based on the selected threshold (distortion level), the intended frequency of the test image is marked as being inside of the display passband (if the distortion to signal ratio is smaller than the threshold) or being outside of the display passband (otherwise). At step five, we group all images with a given disparity level. For each group, we find all frequencies with the distortion to signal ratio smaller than the threshold. We combine these frequencies into a passband region for the given disparity level. In the final step, each passband is approximated by a rectangle. Each of these rectangles has the area and the horizontal to vertical ratio of the corresponding passband region. The horizontal and vertical sizes of that rectangle are used for estimating the equivalent resolution of the display for the corresponding disparity. The output of the last step is a list of equivalent display resolutions for multiple disparity levels and different thresholds. In this paper the analysis is done for two thresholds, 5% and 20%.

Step two of the quality evaluation methodology requires knowledge of the interleaving pattern of the measured display. Note, that it is possible that the interleaving pattern, if provided by the display manufacturer, is a simplified version of the actual one and does not accurately identify the groups of display elements with similar angular visibility functions.

Since the interleaving pattern is an important part in the described quality evaluation method, it has to be known (or evaluated) as correctly as possible. In a previous work, we have described an approach for deriving number of views and interleaving pattern of a given display [12].

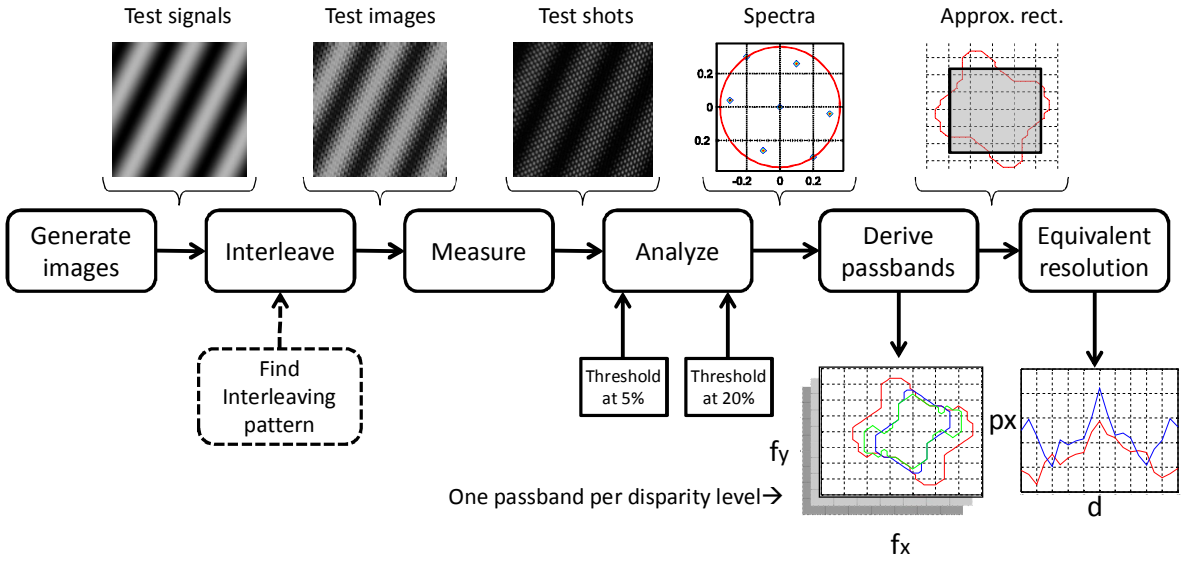


Figure 3 Block diagram of the proposed quality evaluation methodology.

3. MEASUREMENTS

3.1 Generation of test images

The aim of the test image generation procedure is to create a collection of images containing various frequency components, with different apparent depth. The first step is to generate a number of 2D test signals where the brightness of each pixel is calculated by using

$$I = \frac{1}{2}\sin(\pi \cdot x \cdot f_x + \pi \cdot y \cdot f_y) + \frac{1}{2},$$

where I is the brightness, x, y are the horizontal and vertical coordinates of the pixel and f_x, f_y are the horizontal and vertical frequency components, correspondingly. Furthermore, $x = 1,2,...,x_{\max}$, where $x_{\max}$ is the width of the test signal, $y = 1,2,...,y_{\max}$, where $y_{\max}$ is the height of the test signal, $f_x \in [0,1]$ where 1 is normalized to be half the horizontal sampling rate and $f_y \in [-1,1]$ where 1 is normalized to be half the vertical sampling rate. Although it is beneficial to have as many as possible frequency pairs (f_x, f_y), in order to keep the number of test images reasonably small, in our experiments we increase f_x and f_y with step of 0.025. It should be pointed out that in order to generate signals with all possible frequencies that can be displayed we have to use $f_x, f_y \in [-1,1]$. However, due to the symmetrical properties of the 2D discrete Fourier transform (DFT) when applied on real-valued signals it is enough to

use only half of the frequency space as selected above – the magnitude response for a signal with frequency (f_x, f_y) is identical to the one for signal with frequency $(-f_x, -f_y)$ [25].

The next step is to render a number of test images from each test signal, by adding different disparity to each view. First, one should take v copies of the test image, where v equals the total number of views for the measured display and assign them to the views of the display. Then, the contents of each view are shifted horizontally with an offset $s_n = d \cdot n$, where s_n is the offset for the n-th view, n is the view number and d is the targeted disparity. In our experiments d varies between -10 and 10. Finally, the n views are interleaved into a test image, according to the interleaving map of the display. Test images with negative d have apparent depth in front of the screen and test images with positive d have apparent depth further away than the screen plane. Note that in this case *disparity* refers to the disparity between the views, and not the perceived disparity. The former is the offset in pixels between images in neighboring views and the latter is the offset between the images seen by each eye. The artifacts caused by the optical layer are visible by a single eye, and can be measured by a single camera. Such artifacts do not affect the perceived disparity; therefore second camera is not necessary.

In our experiments, we used 23" 3D-display manufactured by X3D-Technnologies, hereafter referred to as *X3D-display*. The display is marketed as an 8-view display, has TFT-LCD matrix with resolution of 1920x1200 and wavelength-selective optical layer which acts as parallax barrier [26]. Since we had a rough estimate of the passband of the display from previous measurements, our test signals did not include all frequency combinations. We prepared 441 tests signals with 21 disparity steps, which resulted in 9261 test images. Three of our test images are shown in Figure 4. Each of them is generated using $f_x = 0.2$ and $f_y = 0.1$, the test image in Figure 4a has disparity $d = 0$ and the test images in Figure 4b and Figure 4c have $d = 1$ and $d = 5$, respectively.

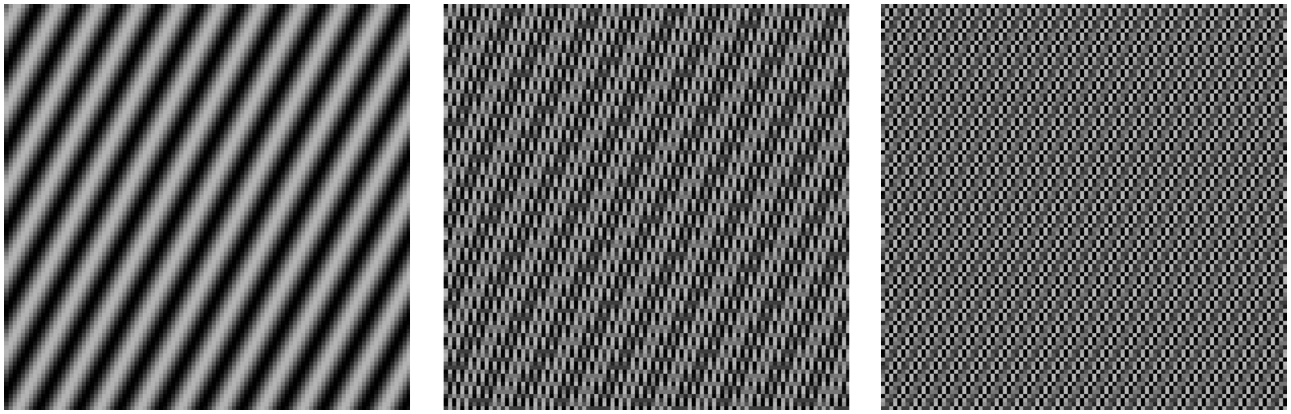


Figure 4 Example of test images with $f_x = 0.2$, $f_y = 0.1$ and different disparity (enlarged details): a) $d = 0$, b) $d = 1$, c) $d = 5$

3.2 Experimental setup

The next step is to visualize each test image on the display and photograph it. In our experiments, we used an HD resolution camera with GigE interface and custom software which automates the visualization-capture-store cycle. The camera was positioned at a distance of 70cm from the display, which is within the nominal observation distance of the X3D-display [26]. In order to avoid aliasing and minimize the influence of the camera, one should use zoom factor that gives the highest possible ratio between the number of photographed display pixels and the number of pixels in the test shot. In our measurement the ratio was 2.24 pixels in the test shot for each pixel in the test image.

The brightness and contrast settings of the display can affect the visibility of image details and thus – the perceptibility of artifacts. Naturally, the visual quality of any display is influenced by its calibration. Our suggestion is to measure the display in typical observation conditions, with values for contrast, brightness and gamma perceptually calibrated to ensure the largest amount of distinguishable levels of gray. In our measurements, the contrast of the display was set to 50%, brightness to 100% and the gamma was set using the visual gamma calibration procedure provided by the drivers of the video card. In order to avoid measurement noise, one should select the lowest ISO sensitivity of the camera and choose exposure time that gives sufficient dynamic range without saturation. We used ISO 50 and exposure time of 1.5sec. The images were captured and converted to gray scale with intensity range between 0 and 217. The three test shots shown in Figure 5 are photographs of the corresponding test images in Figure 4. By comparing images in Figure 4b and Figure 5b, one can see that for the selected f_x and f_y the optical layer of the display works well for $d = 1$, leaving the intended signal predominantly visible. The comparison between images in Figure 4c and Figure 5c, shows that for $d = 5$, the optical layer has undesirable effect on the same combination of f_x and f_y .

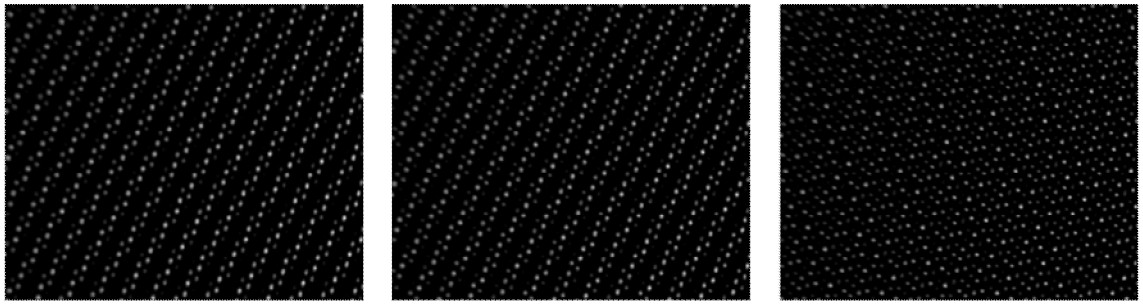


Figure 5 Example of test shots with $f_x = 0.2$, $f_y = 0.1$, and various disparity, acquired during the experiment (enlarged details): a) $d = 0$, b) $d = 1$, c) $d = 5$.

4. DISPLAY PERFORMANCE IN FREQUENCY DOMAIN

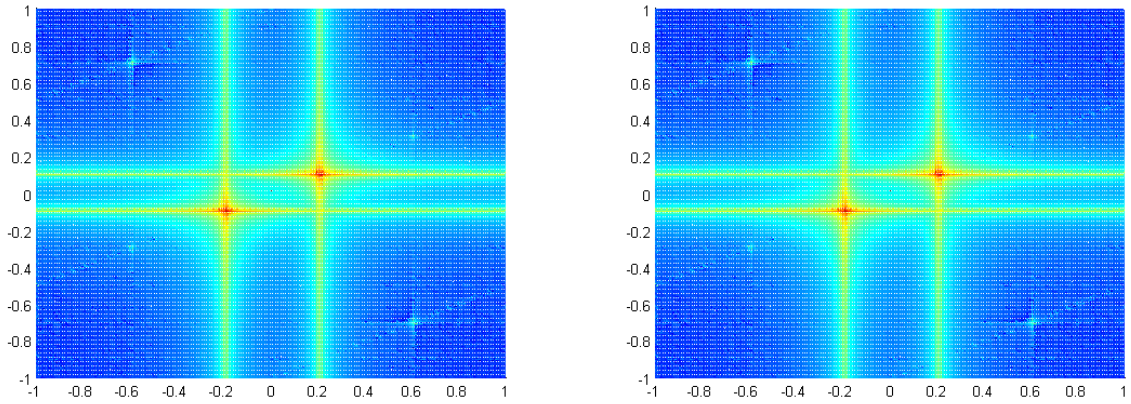
In this section we describe the procedure for evaluating the performance of the display in the frequency domain by processing the test shots obtained as described in the previous section. We do all processing in the frequency domain in order to simplify dealing with various problems that might occur during measurement, for example, camera position, difference in pixel size in the camera and pixels on the display, etc. [12].

4.1 Analysis of frequency components

In an ideal case, a test shot should be visually identical to the test image. However, as discussed in Section 2.1, this is not the case in practice due to the various distortion introduced by a multiview display. For a viewer, when looking at the display, it is obvious if an image is properly represented on the display (see Figure 5a, Figure 5b) and when it is not (see Figure 5c). Such clear identification of visible pattern, as can be done by a human, is not straightforward to obtain by using a signal processing algorithm. Moreover, some of the introduced distortions are more disturbing for a viewer than others. For example, as seen in Figure 5a and Figure 5b, there are many high frequency distortions (dark gaps in the lines). However, HVS is trained to find underlying patterns by grouping features together (*Gestalt principle*) and is sensitive to the predominant frequency components (*pattern masking*) [19]. Therefore, the distortions seen in Figure 5a and Figure 5b, do not hinder the visibility of the original texture. In this figure we still easily see the diagonal lines that we rendered on the display. On the other hand, in the case of low frequency distortions as seen in Figure 5c our original signal is lost. In some cases we will even see signals that did not exist in the test signal but were created by the display and became dominant. Based on the above discussion, we determined criteria in the frequency domain for estimating if

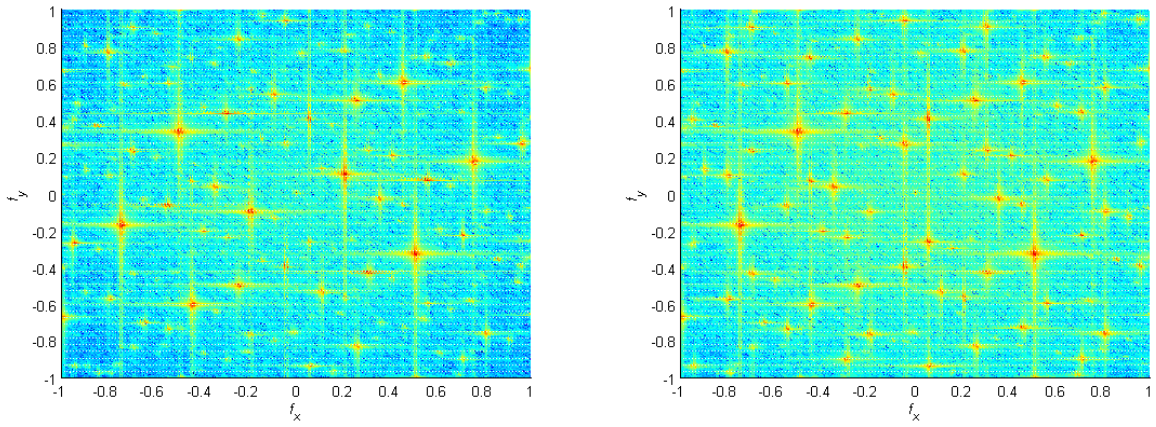
a signal of a particular frequency will or will not be properly represented on the display. The overall procedure implementing the criteria can be summarized in the following four steps[1]:

First, we calculate the spectrum (magnitude of the 2D DFT) of a test shot. Due to various properties of the display, the spectrum of the test shot is very different from the spectrum of the input image. As an example, the spectra for shots shown in Figure 5b and Figure 5c are shown in Figure 6c and Figure 6d, respectively. For comparison, the spectra of the corresponding input signals are shown in Figure 6a and Figure 6b. Since disparity corresponds to shifts in the spatial domain and the magnitude of the DFT is shift invariant, the spectrum (magnitude of DFT) does not depend on the disparity. Therefore, Figure 6a and Figure 6b are identical. On the other side, in both spectra of test shots, Figure 6c and Figure 6d, there are many dominant components. They appear due to the optical effects of the display (optical layer) as it was discussed in Section 2. However, as mentioned before, most of those are high frequency distortions that we can ignore since they will be partially masked by the contrast sensitivity function of the HVS. Moreover, we are not able to do anything about them since they are always present in a multiview display.



a) b)

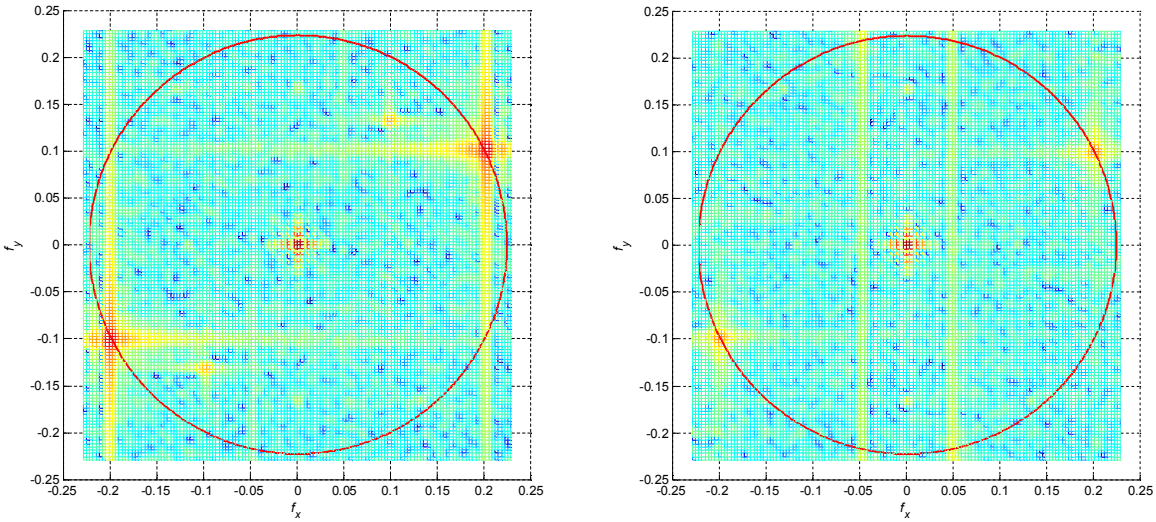
[1] This procedure has been originally introduced in [12] for evaluating the frequency behavior of an autostereoscopic display at zero disparity. Here we will repeat it for completeness together with some additional clarifications and observations



c) d)

Figure 6 Spectra of signal $f_x = 0.2, f_y = 0.1$ at various stages. a) input signal for $d = 1$, a) input signal for $d = 5$, c) test shot for $d = 1$, d) test shot $d = 5$.

Second, based on the observation discussed at the beginning of this section, from the distortion viewpoint, we are only interested in the region containing frequencies lower than the frequency of the input signal. These frequencies lie inside a circle with the center at DC and radius $r_0 = \sqrt{f_{x_0}^2 + f_{y_0}^2}$ with f_{x_0} and f_{y_0} being the frequencies of the input signal in horizontal and vertical direction, respectively. Zoomed detail of the spectra given in Figure 6c and Figure 6d is shown in Figure 7a and Figure 7b, respectively.

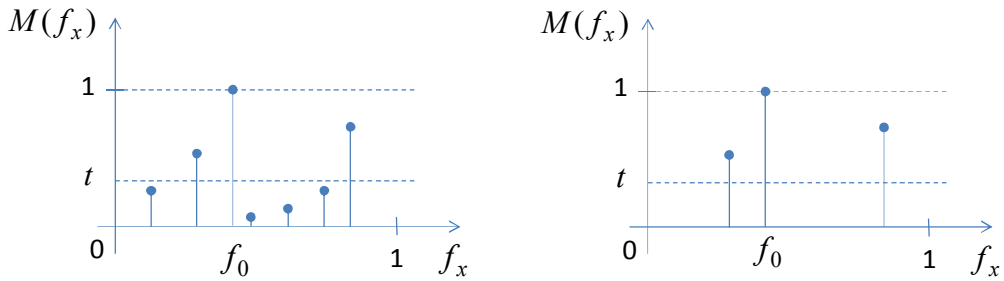


a) b)

Figure 7 Zoomed spectra of test shots for $f_x = 0.2, f_y = 0.1$ in the region of interest. a) $d = 1$, b) $d = 5$.

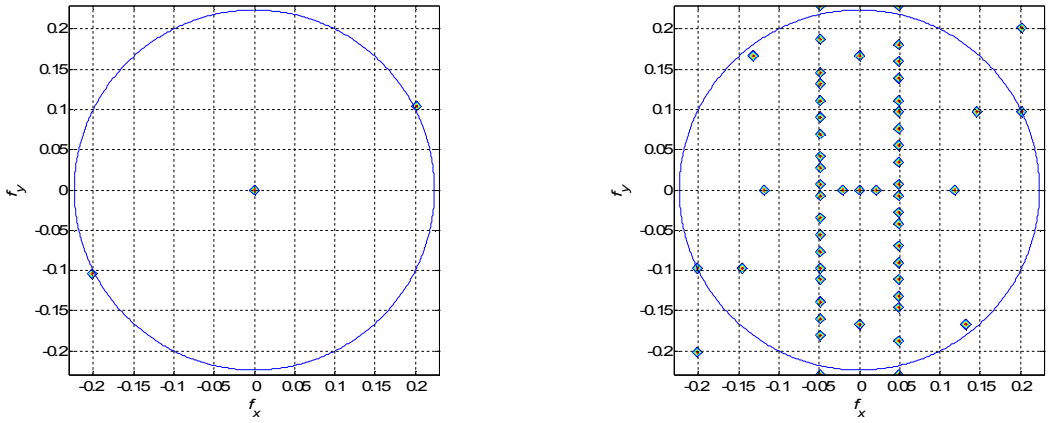
Third, as seen in Figure 7, there are many different signal components present in both test shots. In practice, many of those will not be visible because the amplitude is too small. Therefore, we have to threshold the spectra, that is,

determine when a component is significant and when not. This is directly related to the visibility of various distortions as discussed in Section 2.3. Although, in Section 2.3 the distortion criteria were stated in the time domain, due to the fact that the DFT is a linear transform we can directly apply the same thresholds in the spectral domain. Moreover, if the magnitude of the intended signal is scaled to one, then no additional scaling is required. In the evaluation, we assume that every component that is below the threshold does not contribute to the output signal (will not be visible based on the desired criteria) and therefore we ignore it. This is illustrated by means of a simple 1D example in Figure 8. In this figure, f_x is the sampling frequency in one direction, $M(f_x)$ is the magnitude of the 1D DFT, t is the threshold and f_0 is the frequency of the intended signal with magnitude scaled to 1. After applying the threshold, the original spectra in Figure 8a becomes as shown in Figure 8b. As seen from the figures, all frequency components with magnitude less than t are removed from future analysis. Similarly, after applying the threshold of 5% on the spectra of Figure 7 the thresholded spectra are shown in Figure 9. In this figures, for a better visualization, only the centers of the peaks are shown.



a) b)

Figure 8 A simplified 1D example of tresholding in the spectral domain. a) Spectra before tresholding. b) Spectra after tresholding (all components below the threshold level t have been removed).



a) b)

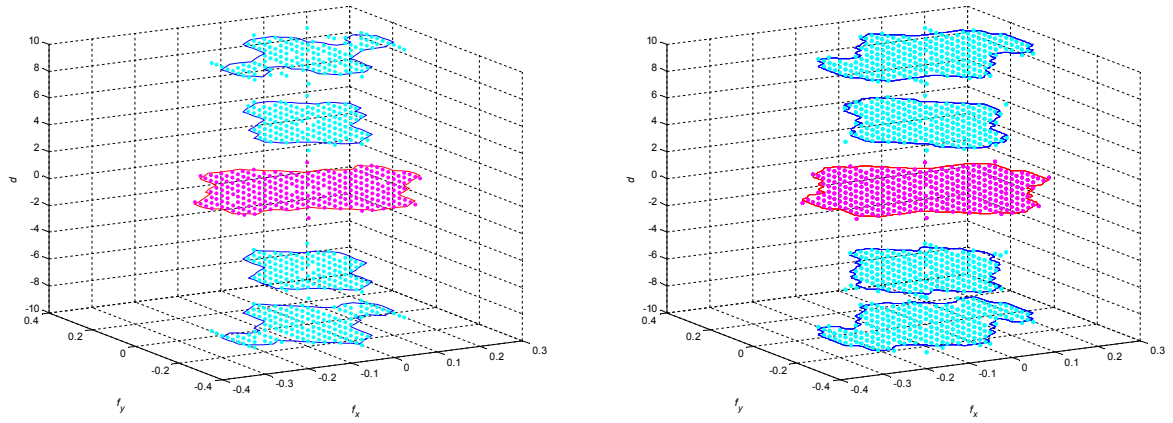
Figure 9 Spectra of test shots for $f_x = 0.2, f_y = 0.1$ in the region of interest represented by the circle after a 5% threshold has been applied. a) $d = 1$, b) $d = 5$.

Fourth, if after applying the threshold there are no signals left with frequencies lower than the input signal, then we assume that signal of this frequency is represented properly on the screen. Consequently, we declare that this frequency is in the passband of the display. This is illustrated in Figure 9a. Since after tresholding, there are not any components left inside the region of interest (marked by circle with radius r_0), this signal ($f_x = 0.2$, $f_y = 0.1$ and $d = 1$) will be properly represented on the display. On the other hand, if there are one or more signal components left, the image on the display will be considerably distorted. Those frequencies we declare as stopband. This is illustrated in Figure 9b. Since after tresholding, there are several components left inside the region of interest, this signal ($f_x = 0.2, f_y = 0.1$ and $d = 5$) will not be properly represented on the display.

4.2 Calculation of display passband

We repeat the above procedure for all shots (for all input frequencies and all disparities). This results in data describing the passband regions at different apparent depths. Furthermore, for each disparity level, we applied a 3x3 median filter in order to smooth the passband region and remove possible gaps caused by non-ideal measurements. The effect of the median filter is rather positive in filling in gaps and the errors it might introduce are negligible with respect to the subsequent approximation of the filter passband. A filter approximating the measured passband region being of reasonable size will always have quite wide transition band, that is, it will be far away from an ideal one and as such the errors introduced by the filter around the edges of the passband will be bigger than the ones introduced by the median filter.

The passband regions for disparities (d = -10, -5, 0, 5 ,10) and thresholds 5% and 20% are given in Figure 10a) and Figure 10b), respectively. The dots show the evaluated data and the solid line around shows the passband edge after median filtering.



a) b)

Figure 10 Display passband for different disparities d = -10, -5, 0, 5 ,10. a) distortion threshold t = 5%, b) distortion treshold$t$ = 20%.

Three observations can be made based on the presented figures. First, the passband form is clearly disparity-dependent. Having the measured pass-bands for different disparities, one can more accurately prepare 3D content to be shown on the display. Second, the passband is dependent on the chosen distortion threshold. The level of 5% corresponds to the visibility threshold and the level of 20% corresponds to a high, but still acceptable, amount of distortions. Thus, measurements at those two levels set up the quantitative compromise between allowing more frequency content to pass versus increasing the amount of visible distortions. In other words, one can design a set of filters ranging from more-restrictive to less-restrictive ones and corresponding to different amount of visible distortions. It can be left to the user's preference to select which filter is to be applied to the watched content. Third, this figure can be used as quality profile. By comparing the passbands of two displays one can judge which of them is better in representing 3D content within given disparity range. The bigger the area of a passband is, and the closer it is to a square, the better suited the display is for visualizing natural content.

5. EQUIVALENT RESOLUTION

In this section we introduce the notation of equivalent resolution. The equivalent resolution is a simplified way to interpret the measured passband for a given threshold and given disparity. Referring to the quality profile of the display given in Figure 10, we note that it might be a bit difficult to use it when comparing this display with other displays. In an attempt to find a simplified yet reasonable representation of the shapes, we approximate the passband for each disparity level with a rectangular shape. The approximating rectangle is centered at the origin, has the same area (in size) as the original passband and overlap as many as possible passband points, while keeping the aspect ratio between maximum values in horizontal and vertical direction. In order to do this, the following set of equations has to be solved:

$$\frac{y_m}{x_m} = \frac{b}{a}$$

$$a \cdot b = A \,,$$

where a and b are the horizontal and vertical width of the rectangle, respectively, x_m and y_m are the maximum width and height of the original shape, respectively, and A is the area of the original shape. These parameters are illustrated in Figure 11. After some trivial mathematical operations, a and b, can be evaluated as follows:

$$a = \sqrt{A\frac{x_m}{y_m}}$$

$$b = \sqrt{A\frac{y_m}{x_m}} \,.$$

As an example, Figure 11 shows the approximation for zero disparity and 5% threshold.

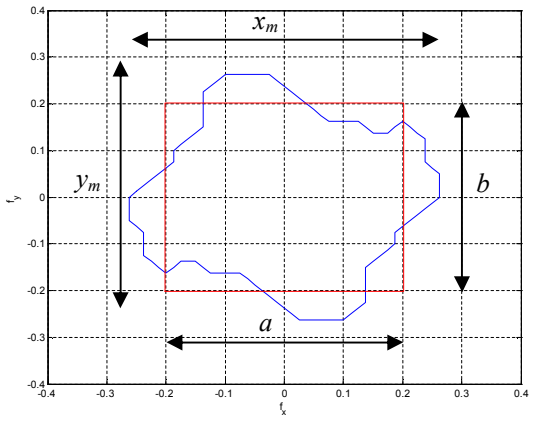
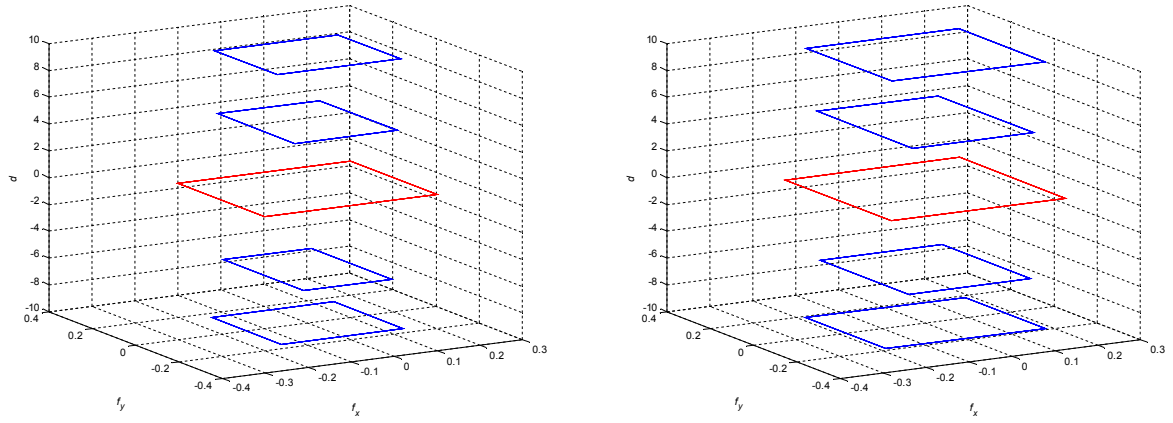


Figure 11 Fitting rectangle to the passband. Example for $t = 5\%$ and $d = 0$.

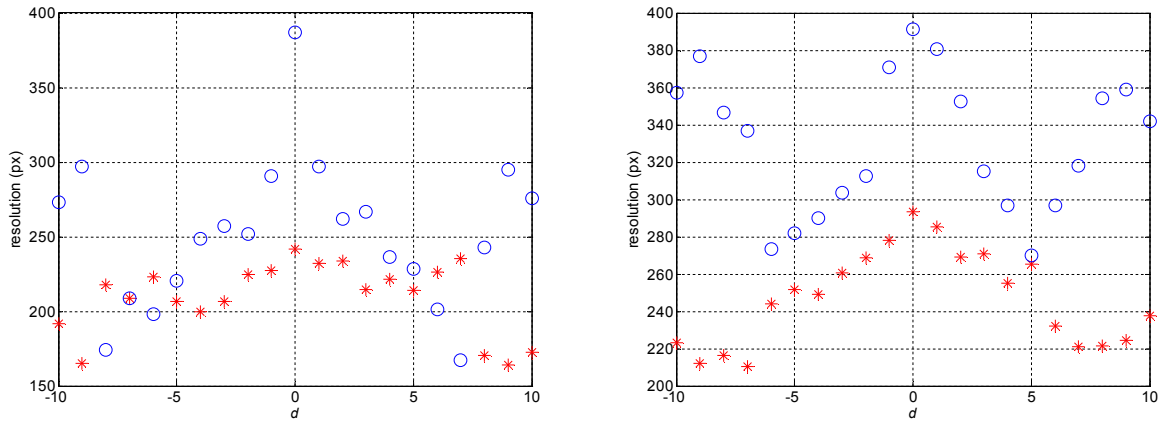
By fitting rectangles for all disparities, for the X3D-display, the equivalent passbands for $t = 5\%$ and $t = 20\%$ are shown, in Figure 12a) and Figure 12b), respectively.



a) b)

Figure 12 Display passbands approximated with rectangles for different disparities d = -10, -5, 0, 5 ,10. a) distortion t = 5%, b) distortion t = 20%.

In order to represent this figure in a more understandable way, we transfer the pasbands in the equivalent resolutions (in number of pixels) in horizontal and vertical direction and plot them with respect to disparity. The equivalent resolution is obtained by multiplying the passband width (height) with the resolution of the display's TFT-LCD matrix in horizontal (vertical) direction. In the case of X3D-display, the TFT-LCD resolution is 1920 by 1200. The equivalent resolution for the X3D-display for the t = 5% and t = 20% is shown in Figure 13a) and Figure 13b), respectively.



a) b)

Figure 13 Equivalent resolution in horizontal (circle) and vertical (star) direction as a function of disparity. a) t = 5%, b) t = 20%.

The equivalent resolution given in Figure 13 is a simplified representation of the true bandwidth in terms of spatial resolution (number of pixels in horizontal and vertical direction for each disparity level). It is calculated to serve two main purposes. First, it enables a fast and easy comparison between different displays. A display that has a higher equivalent resolution at a given disparity will pass more data and as such will be better. Second, the equivalent

resolution is useful when preparing content to be represented on the display. It suggests in an immediate way what the limits in terms of spatial and frequency resolutions are so to avoid preparing images which will be shown improperly on the display. We choose to express the equivalent resolution in pixels, because most users know what visual quality to expect for an image with a resolution given in these units.

6. CONCLUSIONS

In this work, we have drawn a generalized model of a multiview display and used it to explain the reason behind common artifacts, such as aliasing and crosstalk. We have proposed a measurement methodology, which can assess the visibility of these artifacts in patches with different spatial frequency, orientation and disparity. Using these measurements, we have shown how one can derive the display passband for images with different apparent depth. The measurements for display passband versus object disparity can be used for comparing the visual quality of different multiview displays. Additionally, we have given an example about how the display passband can be used to approximate the effective (equivalent) resolution of a multiview display for 3D content with given disparity.

Other comparative studies focus on characterizing the optical quality of a multiview display. In these studies, a large number of parameters of each display are measured and analyzed - e.g. twelve display parameters in [3], six parameters in [8] and four parameters in [7]. Such large variety of parameters allows displays to be characterized in different ways; however it also makes the comparison and choice of a display complex and rather non-intuitive task for display users. In our work, we propose that the display passband is used as (an additional) indicator of perceptual quality of a multiview display. The advantage over other approaches is twofold: first, it is easier to compare two displays - larger and more uniform passband corresponds to a 3D display capable of visualizing a wider range of spatial frequencies; second, it is easier to judge the expected quality for 3D content with given resolution and disparity range - by analyzing the frequency components of a 3D content one can judge if it is suitable for a given display. The measurement results for equivalent resolution versus disparity can be used to optimize content resolution for a given multiview display.

REFERENCES

[1] S. Pastoor, “3D displays”, in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

[2] N. Dodgson, “Autostereoscopic 3D Displays,” Computer , vol.38, no.8, pp. 31- 36, Aug. 2005, IEEE (2005)

[3] M. Salmimaa, T. Jarvenpaa, “Optical characterization of autostereoscopic 3-D displays”, *J. Soc. Inf. Display 16*, 825 (2008)

[4] V. Berkel and J. Clarke, “Characterisation and optimisation of 3D-LCD module design”, *in Proc. SPIE Vol. 2653, Stereoscopic Displays and Virtual Reality Systems IV*, (Fisher, Merritt, Bolas, edts.), p. 179-186, May 1997

[5] J. Konrad and P. Agniel**,** "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Process.*, vol. 15, pp. 128-140, Jan. 2006

[6] V. Saveljev, J.-Y. Son, B. Javidi, S.-K. Kim, and D.-S. Kim, "Moiré Minimization Condition in Three-Dimensional Image Displays," J. Display Technol. 1, 347- (2005)

[7] M. Salmimaa, T. Jarvenpaa, “3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays”, *J. Soc. Inf. Display 16*, 1033 (2008)

[8] P. Boher, T. Leroux, T. Bignon, V. Collomb-Patton, “A new way to characterize auto-stereoscopic 3D displays using Fourier optics instrument”, in *Proc. of SPIE, Stereoscopic displays and applications XX* , 19-21 January 2008, San Jose, California, USA

[9] J. Hakkinen, J. Takatalo, M. Kilpelainen, M. Salmimaa, and G. Nyman, “Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width”, J. Soc. Inf. Display 17, 433 (2009)

[10] F.Kooi, A. Toet, “Visual comfort of binocular and 3D displays”, Displays, Volume 25, Issues 2-3, August 2004, Pages 99-108, ISSN 0141-9382, DOI: 10.1016/j.displa.2004.07.004

[11] S. Pastoor, “Human factors of 3D images: Results of recent research at Heinrich-Hertz-Institut Berlin” in Proceedings of IDW’95, vol. 3D-7,pp. 69-72, 1995.

[12] A. Boev, R. Bregovic, A. Gotchev, “Measuring and modeling per-element angular visibility in multiview displays”, Special issue on 3D displays, Journal of Society for Information Display, vol. 18, no. 9, pp. 686-697, September 2010.

[13] A. Boev, R. Bregovic and Atanas Gotchev, "Design of tuneable anti-aliasing filters for multiview displays", Conference "Stereoscopic Displays and Applications", a part of Electronic Imaging Symposium 2011, San Francisco, CA, USA, January 2011.

[14] A. Jain and J. Konrad, “Crosstalk in automultiscopic 3-D displays: blessing in disguise?”, Stereoscopic Displays and Virtual Reality Systems XIV. Edited by Woods, Andrew J.; Dodgson, Neil A.; Merritt, John O.; Bolas, Mark T.; McDowall, Ian E.. Proceedings of the SPIE, Volume 6490, pp. 649012 (2007).

[15] Christian N. Moller and Adrian R. L. Travis. 2005. Correcting Interperspective Aliasing in Autostereoscopic Displays. IEEE Transactions on Visualization and Computer Graphics 11, 2 (March 2005), 228-236. DOI=10.1109/TVCG.2005.28 http://dx.doi.org/10.1109/TVCG.2005.28

[16] W. IJsselsteijn, P. Seuntiens and L. Meesters, “Human factors of 3D displays”, in (Schreer, Kauff, Sikora, edts.) 3D Video Communication, Wiley, 2005.

[17] A. Boev, R. Bregovic, A. Gotchev, “Methodology for design of antialiasing filters for autostereoscopic displays”, Special issue on Advanced Techniques on Multirate Signal Processing for Digital Information Processing, Journal of IET Signal Processing, to be published (2010)

[18] Shree K. Nayar, Vlad Branzoi, Terry E. Boult, "Programmable Imaging Using a Digital Micromirror Array," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, pp. 436-443, 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) - Volume 1, 2004

[19] B. Wandell, “Foundations of Vision”, Sinauer Associates, 1995

[20] S. Winkler, “Perceptual Video Quality Metrics – A review”, in H. Wu and K. Rao, eds. “Digital video image quality and coding”, ch. 5, CRC press, 2006.

[21] E. Montag and M. Fairchild “Fundamentals of Human Vision and Vision Modelling”, in H. Wu and K. Rao, eds. “Digital video image quality and coding”, ch. 2, CRC press, 2006.

[22] Maria Amparo Díez-Ajenjo, Pascual Capilla, Spatio-temporal Contrast Sensitivity in the Cardinal Directions of the Colour Space. A Review, Journal of Optometry, Volume 3, Issue 1, 2010, Pages 2-19, ISSN 1888-4296, 10.3921/joptom.2010.2.

[23] L. Wang, K. Teunissen, T. Yan, C. Li, Z. Panpan, Z. Tingting, I. Heynderickx, "Crosstalk Evaluation in Stereoscopic Displays," *Display Technology, Journal of* , vol.7, no.4, pp.208-214, April 2011 doi: 10.1109/JDT.2011.2106760

[24] P.J.H. Seuntiëns, L.M.J. Meesters, W.A. IJsselsteijn, “Perceptual attributes of crosstalk in 3D images”, *Displays*, Volume 26, Issues 4-5, October 2005, Pages 177-183, ISSN 0141-9382, 10.1016/j.displa.2005.06.005. (http://www.sciencedirect.com/science/article/pii/S0141938205000429)

[25] R.C. Gonzalez and R.E. Woods, Digital image processing, 3rd. Edition, Prentice Hall, 2007.

[26] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in Proc. SPIE Photonics West 2002: Electronic Imaging, vol. 4660, pp. 212-221, 20023D

[P02] Atanas Gotchev, Gozde Bozdagi Akar, Tolga Capin, Dominik Strohmeier, Atanas Boev, “Three-Dimensional Media for Mobile Devices”, *Proceedings of the IEEE*, April 2011, Vol. 99, 4, pp. 708-737, DOI: 10.1109/JPROC.2010. 2103290

Correction: Word in Fig. 19, second ring, top. Correct: “aberration”.

This page left blank

3D Media for Mobile Devices

Atanas Gotchev, Gozde Bozdagi Akar, Tolga Capin, Dominik Strohmeier, Atanas Boev

Atanas Gotchev (contact author)
Department of Signal Processing, Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland
Phone: +358 3 3115 4349; fax: +358 3 3115 3817
Email: atanas.gotchev@tut.fi

Gozde Bozdagi Akar
Department of Electrical and Electronics Engineering, Middle East Technical University
Ankara, Turkey
Ph: 90 312 2102341; fax: 90 312 2102304
Email: g.bozdagi@ieee.org

Tolga Capin
Bilkent University
Computer Engineering Department
06800 Bilkent, Ankara, Turkey
Phone: +90 312 290 3404; Fax: +90 312 266 4047
Email: tcapin@cs.bilkent.edu.tr

Dominik Strohmeier
Institute for Media Technology, Ilmenau University of Technology
P.O. Box 100565, DE-98684 Ilmenau, Germany
Phone: +49 3677 69 2671; fax: +49 3677 69 1255
Email: dominik.strohmeier@tu-ilmenau.de

Atanas Boev
Department of Signal Processing, Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland
Phone: +358 3 3115 4349; fax: +358 3 3115 3817
Email: atanas.boev@tut.fi

3D Media for Mobile Devices

Atanas Gotchev, Gozde Bozdagi Akar, Tolga Capin, Dominik Strohmeier, Atanas Boev

ABSTRACT

This paper aims at providing an overview of the core technologies enabling the delivery of 3D Media to next-generation mobile devices. To succeed in the design of the corresponding system, a profound knowledge about the human visual system and the visual cues which form the perception of depth, combined with understanding of the user requirements for designing user experience for mobile 3D media, are required. These aspects are addressed first and related with the critical parts of the generic system within a novel user-centered research framework. Next-generation mobile devices are characterized through their portable 3D displays, as those are considered critical for enabling a genuine 3D experience on mobiles. Quality of 3D content is emphasized as the most important factor for the adoption of the new technology. Quality is characterized through the most typical, 3D-specific visual artifacts on portable 3D displays and through subjective tests addressing the acceptance and satisfaction of different 3D video representation, coding and transmission methods. An emphasis is put on 3D video broadcast over DVB-H in order to illustrate the importance of the joint source-channel optimization of 3D video for its efficient compression and robust transmission over error-prone channels. The comparative results obtained identify the best coding and transmission approaches and enlighten the interaction between video quality and depth perception along with the influence of the context of media use. Finally, the paper speculates on the role and place of 3D multimedia mobile devices in the future internet continuum involving the users in co-creation and refining of rich 3D media content.

Keywords: 3D visual artifacts, auto-stereoscopic displays, graphical user interface, multi-view coding, MPE-FEC, open profiling of quality, user-centric design

I. INTRODUCTION

3D media is an emerging set of technologies and related content in the area of audio-video entertainment and multimedia. It is expected to bring realistic presentation of third dimension of audio and video and to offer immersive experience to the users consuming such content. While emerging in areas such as 3D cinema and 3D television, 3D media has also been actively researched for its delivery to mobile devices.

The general concept of 3D media assumes that the content is to be viewed on big screens and simultaneously by multiple users. Glasses-enabled stereoscopic display technologies have matured sufficiently to back the success of 3D cinema and have also been enabling the introduction of first generation 3DTV. Autostereoscopic displays have been developed as an alternative display technology offering glasses-free 3D experience for the next generation 3DTV. Advanced light-field and holographic displays have been anticipated in the mid-term future. On the research side, various aspects of 3D content creation, coding, delivery and system integration have been addressed by numerous projects and standardization activities [1], [2], [3]. At a first sight, these developments position 3D Media as a rather diverging technology with respect to mobile multimedia as the former relies on big screens and realistic visualization and the latter relies on portable displays. Still, a symbiosis between 3D and mobile media has been considered rather attractive. 3D would benefit from being introduced also to the more dynamic and novel technology-receptive mobile tech market. Mobile TV and video and the corresponding broadcasting standards would benefit from the rich content leading to new business models. The research challenge of achieving this symbiosis is to adapt, modify and advance the 3D video technology, originally targeted for large screen experience, for the small displays of handhelds.

The introduction of 3D media to handhelds is supported by the current trend of developing novel multicore processors as an effective way to reduce the power consumption while maintaining or increasing the performance [4]. Increasing the number of cores and thus offering parallel engines is perfectly suitable for 3D data, which naturally call for parallel processing. New multicore platforms for mobile applications offer balanced architectures to support both data-dominated and control-dominated applications [5]. Examples are the Texas instruments' OMAP 4 [6], NXP's LH7A400 [7], Marvell's PXA320 [8], NVIDIA Tegra APX 2500/2600 Series, Next Generation NVIDIA Tegra [9], [10], Qualcomm Snapdragon Series [11], and ST Ericsson's U8500 [145]. The aim in designing such multicore processors has been to achieve high system clock rate, optimize the memory use and interconnections between cores and provide functionality for new rich multimedia applications by more powerful graphical accelerators and digital signal processors. Support of 3D graphics for 3D user interfaces and 3D gaming as well as existing and future multimedia encoders has been targeted. Specifically, 3D rendering has been considered to be implemented primary on a dedicated hardware accelerator than on a general-purpose CPU, allowing both faster execution and lower power consumption,

which are crucial for mobile devices. In addition, modern APIs, such as OpenGL ES 2.0, emphasize parallel processing design, making it also possible to support more advanced and data-intensive 3D applications on a mobile device. One of the research challenges is to design efficient 3D processing algorithms, which reduce the internal traffic between the processing elements and the memory, while maintaining low power consumption [12]. While modern multicore development platforms are available for integrating 3D video decoding, processing and playing algorithms, it is the new portable 3D displays which should make the difference in delivering new user experience.

This paper analyses the process of bringing 3D media to mobiles. Section I analyzes what is important to know *before* beginning the design of a 3D media system for mobiles. The section starts with a brief overview of the basics of depth perception by the human visual system (HVS) and the relative importance of various 3D visual cues. Along with psychophysical factors, novel user studies are presented which help to understand the user expectations and requirements concerning 3D Media for mobiles. The introduction of new media requires also novel research approaches regarding users and new, user-centric, approaches in designing critical parts of the whole system. Those are presented next, just before the overview of the 3D video delivery chain with its main blocks. Emphasizing 3D video is important, as it illustrates the entertainment value of 3D for mobile users. Optimal content formats and coding approaches, as well as streaming and channel coding approaches especially tailored to 3D are reviewed as to make a link to the other papers in this special issue. Thus, Section II connects the user with the system through psychophysical and psychological aspects and the ways those have to be investigated.

Section III is all devoted to portable 3D displays, as the main part of the next-generation 3D-enabled mobile devices playing a decisive role in the adoption of the new technology. Related display technologies are overviewed. Display optical parameters which determine the quality of 3D perception are summarized and measurement results are presented to characterize and compare various displays.

The knowledge about portable 3D displays forms the basics to proceed further with Section IV, where user experience of 3D mobile media is explored in details. 3D-specific artifacts are reviewed and put against the stages of the delivery chain being responsible for their generation and to the specifics of the human visual system. Furthermore, novel studies aimed at identifying best accepted 3D video representation formats, and source and channel coding methods are presented. Objective comparisons are complemented by results from extensive subjective tests based on novel design methodologies. The studies on 3D video are completed at the end of the section with an overview of recent advances in 3D graphical user interfaces.

Section V presents a foreseeing of more futuristic usage scenarios of 3D-enabled handhelds where 3D media is not only *delivered* to users but also *co-created* by them using the tools as envisaged by Future Internet. Such concept poses even

more challenging research questions addressing the way 3D audio and video content is captured and processed by mobiles to contribute to a collaborative creation of rich 3D media content and corresponding services.

II. INTERDISCIPLINARY ASPECTS OF 3D MOBILE MEDIA SYSTEM DESIGN

A. Perception of depth

The human visual system can be considered as a set of separate sub-systems operating together in a unified manner. There are largely independent neural paths responsible for transmitting the spatial, color and motion information to the brain [28]. On perceptual level there are separate visual mechanisms and neural paths, while on cognitive level there are separate depth cues contributing to the formation of 3D spatial vision [28], [29]. These depth cues are with varying importance for an individual observer [30], [31], [32]. The depth cues used for assessing the depth by different layers in human vision are shown in Figure 1 and are as follows:

- Accommodation – This is the ability of the eye to optically focus on objects at various distances.
- Binocular depth cues – These result from the position of the two eyes observing the scene from slightly different angles. The eyes tend to take a position which minimizes the difference of the visual information projected in both retinae. The process is called *vergence* and can be characterized by the angle between the eyes used as a depth cue. With the eyes converged on a point, *stereopsis* is the subsequent process which uses the residual disparity of the surrounding area for depth estimation relative to the point of convergence.
- Pictorial cues – These include shadows, perspective lines, texture scaling and can be perceived even with a single eye.
- Motion parallax – this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. Similar mechanism has been observed to be used by insects and is commonly referred to as “insect navigation” [38].

A 3D media system has to maintain adequate 3D visual cues. Accommodation is the primary depth cue for very short distances, where an object is hardly visible with two eyes. Its importance decreases sharply with increasing the distance. HVS tends to combine accommodation with convergence, using the information from the latter to correct the refraction power and to ensure clear image of the object being tracked. In the real world accommodation and convergence points coincides, however on stereoscopic displays they may differ as eyes focus on the screen and try to converge according to the binocular difference. This discrepancy leads to so-called “accommodation-convergence rivalry”, which is a major limiting factor for such displays. Binocular depth cues have been the most used in “3D cinema”, and subsequently in 3D TV and 3D for mobiles, by presenting different-perspective images to the two eyes. Binocular vision is quite vulnerable

to artefacts: an “unnatural” stereo-pair presented to the eyes can lead to nausea and “simulator sickness”, as the HVS is not prepared to handle such information [37]. About 5% of all people are “stereoscopically latent” and have difficulties assessing binocular depth cues [28], [29]. Such people perceive depth, relying only on depth cues coming from other visual layers. Pictorial cues work for longer distances, where binocular depth cues become less important. At medium distances, pictorial and binocular cues are combined and for such distance the perception can be ruined by missing subtle pictorial details, even if stereoscopy is well presented. It is said that the scene exhibits “puppet theatre” or “cardboard effect” artifacts. The motion parallax depth cues might be affected primarily by artifacts appearing in temporal domain such as motion blur and display persistence.

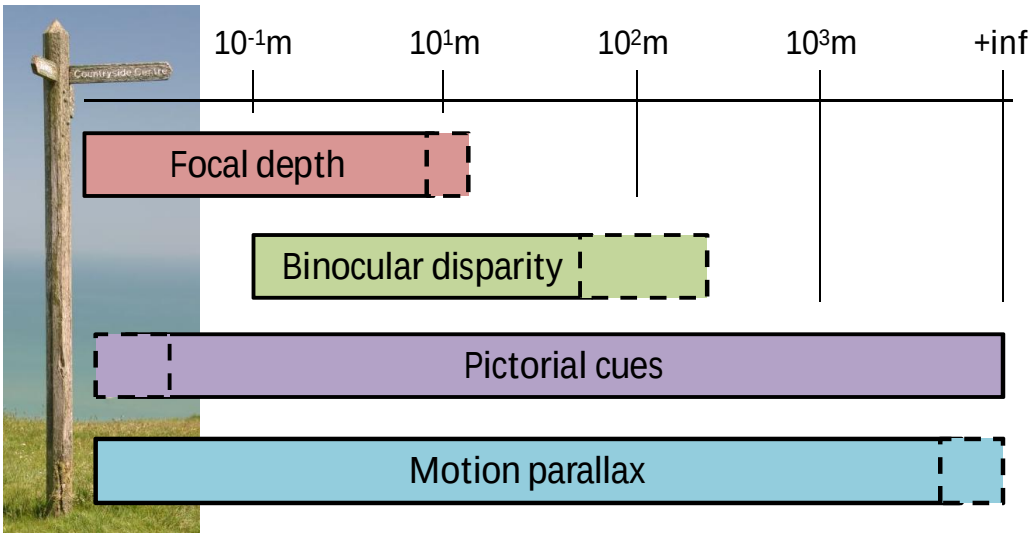


Figure 1. Depth perception as a set of separate visual “layers” (photo by Ard vd Leeuw).

An interesting suggestion is that binocular and monocular depth cues are independently perceived. It has been supported by both subjective experiments (e.g. the famous experiments with so-called “random dot stereograms” [33]) and anatomical findings. The latter have shown that first cells that react to a stimulus presented to either of the eyes (binocular cells) appear at a late stage of the visual pathways, more specifically in the V1 area of brain cortex. At this stage, only the information extracted separately for each eye, is available to the brain for deduction of image disparity [28]. A practical implication of the above suggestion concerns the modeling, assessment and mitigation of visual artifacts building on the hypothesis that “2D” (monoscopic) and “3D” (stereoscopic) artifacts would be perceived independently [34]. Planar “2D” artifacts, such as noise, ringing, etc, are thoroughly studied in the literature [35], [36], while artifacts which affect stereoscopic perception have been addressed more recently [39]. We present more details on 3D visual artifacts in Section IV, after presenting the main blocks of a 3D media system and the specifics of portable 3D displays.

B. User issues at the beginning of 3D media system design

The perception of depth is an important aspect in the development of 3D media on mobile devices. However, an optimized development of such systems must take into account further requirements. Like in every product development process, the goal is that the prospective end product as a whole will satisfy the end users. This satisfaction is a key requirement for the success of the product. To describe users' needs and expectations about the product under development, user requirements are commonly specified before and verified, and if necessary redefined, cyclically during the development process [118]. By definition, user requirements describe any externally visible function, constraint, or other property that a product must provide to reach user satisfaction [139]. However, this product-oriented definition is limited as it overlooks the characteristics of the end users. User experience (UX) tries to understand end users' needs, concerns, and expectations more broadly. It has been defined as being about technology that fulfils more than just instrumental needs in a way that acknowledges its use as a subjective, situated, complex and dynamic encounter [43]. According to Hassenzahl and Tractinsky [43], UX is "*a consequence of a user's internal state […], characteristics of designed system […] and the context […] within the interaction occurs*".

User requirements for designing user experience for mobile 3D media

In the development of 3D media systems and services, the identification of user requirements plays a crucial role. 3D mobile media combines the technologies of 3D media and mobile devices. Each of these technologies has its own user requirements that need to be fused into a new system providing a seamless UX. Mobile media research has identified three building blocks for UX. Roto [44] describes them as (1) user, (2) system and services, and (3) context. Following these building blocks of mobile UX, a large study of a methodological triangulation has been conducted to target the explicit and implicit user requirements for mobile 3D video [116], [117]. In that study, an online survey, focus groups, and a probe study are combined to be able to holistically elicit user requirements. The survey has been used first to identify and verify needs and practices towards the new system. It has been then extended with the results of focus groups. The focus group studies have been conducted to overcome the weakness of online surveys to generate new ideas. More specifically, focus groups aimed at collecting possible use scenarios for mobile 3D media as well as an imaginary design of the device and the relating services. However, both online survey and focus groups only cover the *explicit* user requirements. Especially focus groups do not take into account individual, implicit requirements as those are often overwhelmed by the group effect. To complete the user requirements, the probe study as the third method has been applied to collect those personal needs and concerns. In this probe study, test participants played with a probe package that contained a disposable camera, a small booklet and material for a collage, as illustrated in Figure 2. Their task was to log their thoughts and ideas about mobile 3D media in different daily situations and therewith in different

contexts with help of the diary and the disposable camera. At the end, test participants set up a collage in a reflective task about their own opinion on mobile 3D video. Examples are shown in Figure 3 [116], [117].

Figure 2. Probe package provided to participants during user requirement elicitation studies [117] .

Figure 3. Examples of two collages from probe study on user requirements for mobile 3DTV [117].

The above referred studies [116], [117] have framed the user requirements for mobile 3D video with respect to all three building blocks of UX: the user, the system and service, and the context. The results show that the prospective users of mobile 3D television and video systems want to satisfy *entertainment* and *information* needs. Participants outline the

importance of the added value given through *increased realism* and a *closer emotional relation* to the content. It is noteworthy that these expectations about added value differ from the common ideas about added value of 3D. For large screens or immersive environments, the added value is commonly expressed as *presence*, the users' feeling of being there [119]. Related to system and services, users expect devices with a display of the size of 3-5". The display must provide possibilities to shift content-dependently between monoscopic and stereoscopic presentation. The expected content relates to the entertainment and information needs. TV contents like sports, documentaries, or even news are mentioned by the test participants. However, the requirements also show that non-television content has high potential for the services. Applications like interactive navigation or games are of high interest for the users. To access the different services, users can image both on-demand and push services that will be paid by monthly payment or pay-per-view. The expected use (the context) is mainly in public transports, cafes, or waiting situations and in private viewing, when concentrating on the content. Especially young people have told also about a need for shared viewing. However, interaction with the context (as e.g. defined in Sub-section IV.C) or with other users on one display is not expected regularly. As mobile 3D media is well-suited for waiting situations and short transport trips, the expected viewing time is about 15 minutes. In exceptional cases like journeys also longer duration up to half an hour may occur.

A holistic user-centered research framework for mobile 3D television and video

The elicited user requirements for mobile 3D video show what people expect from this new technology. A challenge during the development process is now how to include these requirements into the technology. The user-centered design process is defined in ISO 13407 [118] as a cyclic process within a product development, as exemplified in Figure 4. It is especially useful at an early stage of the development as it can show opportunities to improve the quality of the system related to the requirements of the prospective end users. However, user-centered design can be used during the whole development process.

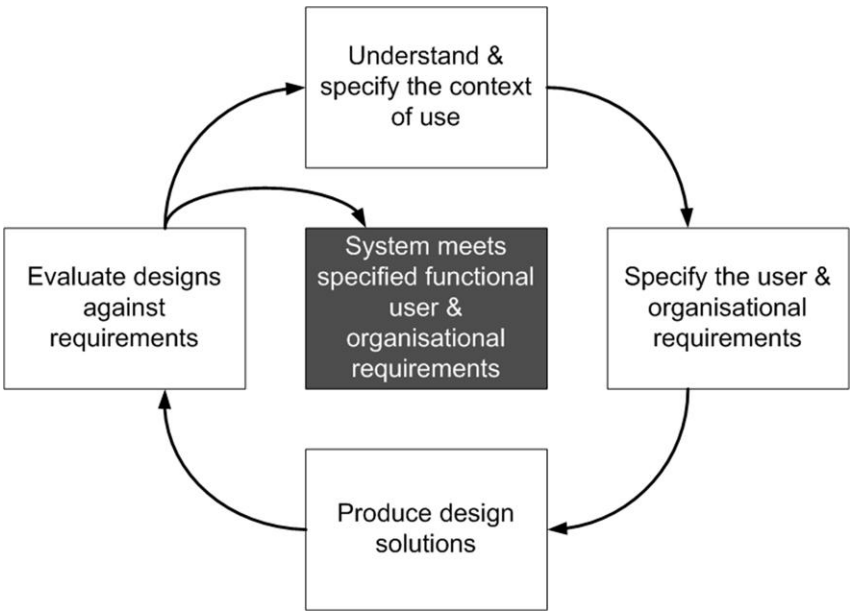


Figure 4. Cyclic process of user-centered design according to ISO 13407 [118].

Current work on mobile 3D media has been conducted under the framework of User-centered Quality of Experience (UC-QoE) [106], [108]. In general, QoE is defined as "the overall acceptability of an application or service, as perceived subjectively by the end-user" [129]. Quality of Experience takes into account the cognitive processes of human perception that relate to interpretation of perceived stimuli with regard to emotions, knowledge and motivation. More broadly, QoE can be regarded as a "multidimensional construct of user perception and behavior" [132]. The UC-QoE approach represents a holistic framework for subjective quality optimization of mobile 3D video. It takes into account prospective users and their requirements, evaluation of system characteristics, and evaluation of quality in the actual context of use [108]. The framework provides a set of evaluation methods to be able to study the different aspects of Quality of Experience. Especially two challenges have been identified along with shortcomings of currently existing quality evaluation methods. Commonly, subjective quality is measured using psychoperceptual evaluation methods that are provided mainly in ITU recommendations [114], [115] (see [106] for a review). First, these methods target a quantitative analysis of the excellence of overall quality disregarding users' quality interpretations, descriptions and evaluation criteria that underline a quantitative quality preference. Second, these methods have been designed for quality evaluations in controlled, homogenous environments. However, mobile applications are meant specifically for use in extremely heterogeneous environments as the user requirements show [109], [116]. To get a higher external validity of the results, these systems must be evaluated additionally in their actual context of use.

There has been a gap between quantitative evaluation of the user satisfaction with the overall quality and the underlying components of quality in multimedia quality evaluation [123]. To address this gap, an approach referred to as *Open Profiling of Quality* (OPQ) has been developed and successfully applied in mobile 3D media research [120], [121], [123]. Open Profiling of Quality is a mixed method that combines evaluation of quality preferences and the elicitation

of individual experienced quality factors [123]. *Sensory profiling*, originally used in food sciences as a research method "*to evoke, measure, analyze and interpret reactions to those characteristics of food and materials as they are perceived by senses of light, smell, taste, touch and hearing...*" [125] has been adapted for 3D media studies. Final outcome of OPQ is a combination of quantitative and sensory data sets connecting users' quality preferences with perceptual quality factors. In its sensory profiling task, test participants develop their own idiosyncratic quality attributes. These attributes are then used to evaluate overall quality [122]. The sensory data can be analyzed using multivariate analysis methods [113], [130] and the results show a perceptual model of the experienced quality factors.

To overcome the limitations of a controlled laboratory environment, the second evaluation tool within the UC-QoE framework is a hybrid method for quality evaluation in the Context of Use [131]. Context of use is defined as the entity of physical and temporal contexts, task and social contexts as well as technical and informational contexts [107], [131]. The extension of quality evaluation to the context of use aims at extending the external validity of results gained in controlled environments. Concrete results of applying the two evaluation tools to characterize UC-QoE of mobile 3D media are given in Section IV.B.

C. 3D media delivery chain for mobiles

A system for delivery of 3D media to mobile devices is conceptualized in Figure 5. On a general level, its building blocks do not differ much from the blocks of a general 3DTV system. The system includes stages of content creation, format conversion to a compression- and delivery-friendly format, compression with subsequent transmission over some wireless channel, decoding and displaying on a mobile terminal.

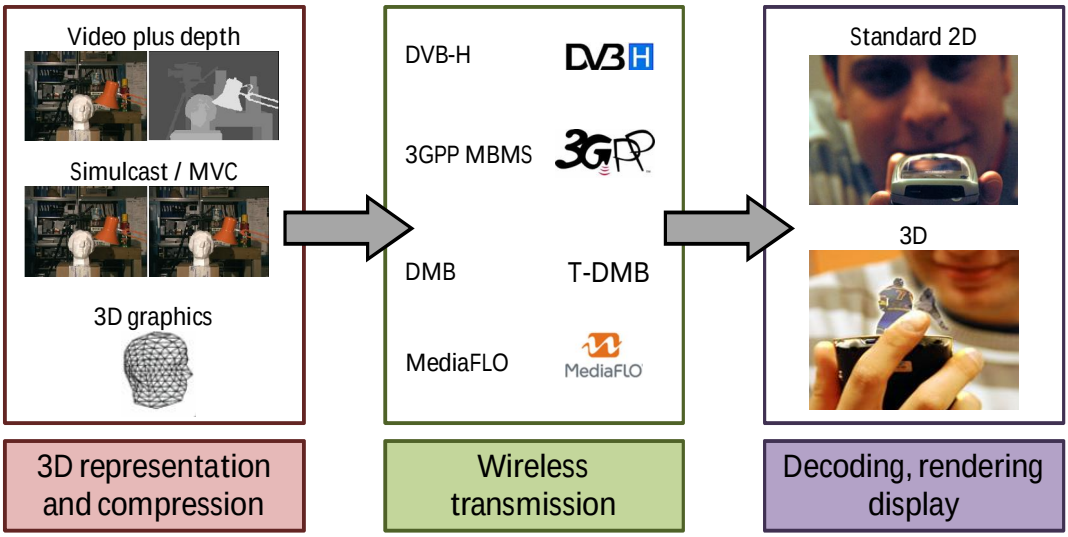


Figure 5. End-to-end 3D video transmission chain.

The specifics of this general system are determined by the foreseen mobile applications such as video conferencing, online interactive gaming, and mobile 3DTV; the characteristics of the wireless networks such as DVB-H, DMB, MediaFlo, 3G and the computational power of the terminal device. For real-time video communication such as video conferencing, real-time encoding and decoding is necessary simultaneously at both terminal devices with low delay. The transmission bandwidth is restricted to the capabilities of the mobile phone line which makes the bitrate for the 3D video signal very limited. For mobile 3DTV, the decoding is only done at the receiver side with some possible buffering. However, in this case, rendering and display at full frame rate and with minimum artifacts is needed. In addition, due to the characteristics of the wireless channel, the quality cannot be guaranteed which brings the necessity of robustness to channel errors. For online interactive gaming, again fluent decoding, rendering and possible content adaptation is needed at the terminal devices with low delay. In addition to all these specific requirements and limitations, low power consumption and low complexity is a must for mobile video applications.

3D video representation and coding

Considering the above limitations, the first issue to look at is the format to be used for the delivery of 3D video and 3D graphics. If the latter is to be transmitted as a polygon mesh, formed by collection of vertices and polygons to define the shape of an object in 3D, then MPEG4 AFX is a well known compression method to be used. 3D video offers more diverse alternatives for its representation and coding and we will concentrate on these other than the 3D graphics. The first research attempts and related standardization efforts regard 3D video represented either by single video channel augmented by depth information (view+depth (V+D)) or by parallel video streams coming from synchronous cameras. In the latter representation approach, the video streams can be compressed jointly (multi-view) or independently (simulcast).

V+D coding: ISO/IEC 23002-3 Auxiliary Video Data Representations (MPEG-C part 3) is meant for applications where 3D video is represented in the format of single view + associated depth (V+D), where the single channel video is augmented by the per-pixel depth attached as auxiliary data [135]. The presence of depth data allows for synthesizing desired views at the receiver side and adjusting the view parallax, which is beneficial for applications where the display size might vary, which is the case of mobile devices. V+D coding does not require any specific coding algorithms. It is only necessary to specify high-level syntax that allows a decoder to interpret two incoming video streams correctly as color and depth. Additionally, it is backward compatible and its compression efficiency is high as the side depth channel is represented by a gray-scale image sequence. Few studies have reported algorithms and prototypes for view synthesis based on V+D (ISO/IEC 23002-3) on mobile devices [142], [143]. Contrary to their compression efficiency, such systems have high complexity for both sender and receiver sides. Before encoding, the depth data has to be precisely generated. For real scenes, this is done by depth/disparity estimation from captured stereo or multi-camera

videos using extensive computer vision algorithms plus possibly involving range sensors. For synthetic scenes, this is done by converting the z-buffer data resulting from rendering based on 3D models. V+D representation is only capable of rendering a limited depth range and additional tools are needed to handle occlusions. Recent advances to this approach suggest using so-called depth-enhanced stereo or multi-layer depth [86], which successfully tackle the occlusion issue for the price of increased complexity. At the receiver side, view synthesis has to be performed after decoding to generate the stereo pair which is not very trivial for mobile devices to achieve in real time especially for high resolutions.

Multiview Video Coding (MVC, ISO/IEC 14496-10:2008 Amendment 1 ITU-T H.264): It is an extension of the Advanced Video Coding (AVC) standard [134]. It targets coding of video captured by multiple cameras. The video representation format is based on *N* views. MVC exploits temporal and inter-view redundancy by interleaving camera views and coding in a hierarchical manner. There are two profiles currently defined by MVC: Multiview High profile and Stereo High profile which are both based on the ITU-T H.264 Advanced Video Coding (AVC) with a few differences [88]. Stereo High profile is also chosen as the supported format for the 3D Blu-ray discs. The main prediction structure of MVC is quite complex introducing a lot of dependencies between images and views. In order to decrease the complexity, an alternative simplified structure is presented in [101] and shown to be very close to the main prediction structure in terms of overall coding efficiency. In this simplified prediction structure, the temporal prediction remains unchanged when compared to original MVC prediction structure, but spatial references are only limited to anchor frames, such that spatial references are only allowed at the beginning of a group of pictures (GOP) between I and P pictures. This simplified version is shown in Figure 6 for stereoscopic video where only two views (left and right views- S_0 and S_1) exist.

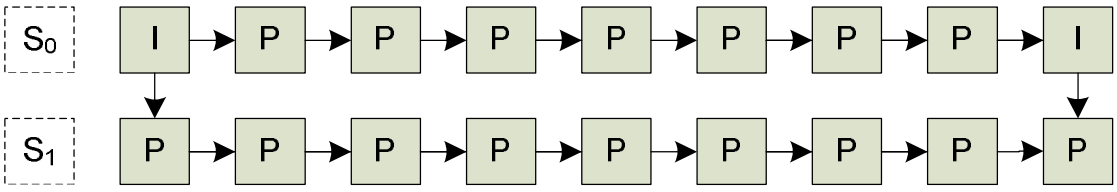


Figure 6. Simplified IPP… prediction structure of MVC codec with inter-view references in anchor frames.

It should be emphasized that this coding is also backward-compatible meaning that the only mono-capable receivers will still be able to decode and watch left view, which is nothing but a 2D conventional video, and simply discard the other view, since left view is encoded independent of the right view.

Research on coding of multi-view video and view plus depth has reached a good level of maturity and the related international standards are perfectly applicable for mobile 3D video systems and services. However, there are inferior

points which prompt for further research. While the approach based on coding of single view plus dense depth seems to be preferred for its scalability, it might be too computationally demanding for the terminal device as it requires view rendering and hence make the device less power efficient. MVC, i.e. compressing the two views by joint temporal and disparity prediction techniques is not always efficient for compression. Researchers have hypothesized that in a mobile device the stereo perception can be based on reduced cues and suggested approaches based on reduced spatial resolution, so-called mixed resolution stereo coding (MRSC) [127]. In this approach, one of the views is kept intact while the other is properly spatially decimated to a suitable resolution where the stereo is still well perceived [127]. Though subjective studies have not proved the MRSC coding hypothesis and such compression has been evaluated inferior to MVC and V+D [122], the approach bears a research potential especially when combined also with MVC-type of motion/disparity prediction [128].

Simulcast coding/Interleaved coding: Another way to code 3D video is to use existing video codecs to stereoscopic video with/without an interleaving approach. If no interleaving is used, one achieves **simulcast coding** that is not any different than coding a conventional 2D video with a video encoder in the sense that both of the views are coded as two completely independent 2D videos [87]. This method allocates the highest bitrate for a video compared to the other solutions, but is the least complex. On the other hand **interleaving** [89] can be used as (a) time multiplexing, (b) spatial multiplexing as over/under, (c) spatial multiplexing as side-by-side as shown in Figure 7 ((b) and (c) are also called as frame-compatible modes). This method is currently used by the broadcasters doing initial 3D trials since both the encoding and decoding can be done with any existing equipment. The losses of either temporal or spatial resolution as well as the reduced robustness to errors position this kind of representation as an inferior with respect the other 3D video representation approaches.

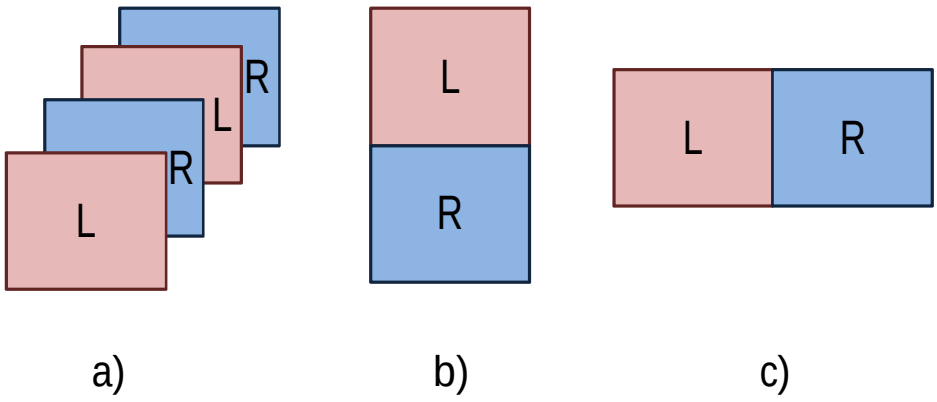


Figure 7. Interleaving of left and right channels (a) Time multiplexing, (b) Spatial multiplexing (up-down), (c) Spatial multiplexing (side-by-side).

Recent activities of the 3DV Video group at MPEG have been focused on combining the benefits of V+D and MVC in a new 3D video coding format so to allow for efficient compression and rendering of multiple views on various auto-stereoscopic displays [144]. Extensions denoted as '*depth-enhanced stereo*' and '*multi-view multi-depth*' have been considered (as also described in this special issue).

Wireless channels

After the coding format selection, the next issue to investigate is the channels to be used for delivery of 3D video to mobile devices. The delivery channels to be used depend heavily on the targeted application. Video-on demand services, both for news and for entertainment applications, are already being offered over the Internet which can be extended to 3D. Also, 3G and 4G mobile network operators use IP successfully to offer wireless video services.

On the other hand, when the same video needs to be distributed to many users, collaboration between the users may significantly enhance the overall network performance. Peer-to-peer (P2P) streaming refers to methods where each user allocates some of its resources to forward received streams to other users; hence, each receiving user acts partly as a sending user.

At the same time, Mobile TV has recently received a lot of attention worldwide with the advances in broadcasting technologies such as Digital Multimedia Broadcasting (DMB), Digital Video Broadcasting - Handheld (DVB-H) and MediaFLO [90] from one side and the 3GPP's multimedia broadcast and multicast services (MBMS) [141] from another.

Currently, there are a number of projects conducting research on transmitting 3D video over such existing infrastructures such as the Korean 3D T-DMB [91], the European 3D Phone [92] , Mobile3DTV [93] addressing the delivery of 3DTV to mobile users over DVB-H system and DIOMEDES [94] addressing 3D Peer-to-Peer (P2P) distribution and broadcasting systems . Recently, DVB has also established 3D TV group (CM-3DTV) to identify "what kind of 3D-TV solution does the market want and need, and how can DVB play an active part in the creation of that solution?" [98].

As summarized in this section, there is a significant amount of work done in the various standards organizations in the area of representation, coding and transmission of 3D data. The most critical part is to find the optimized solution to deliver content with satisfactory quality and give the user a realistic 3D viewing experience on a 3D portable display. These issues will be addressed in the subsequent sections.

III. PORTABLE 3D DISPLAYS

3D display is the most critical part of a 3D-enabled mobile device. It is expected to create lively and realistic 3D sensation, meeting at the same time quite harsh limitations of screen size, spatial resolution, CPU power and battery life. Among the wide range of state-of-the-art 3D display technologies [13], not all are appropriate for mobile use. For mobile phones or personal media players, wearing glasses or head-mounted displays to aid the 3D perception would be rather inconvenient. Volumetric and holographic displays are far from mature for mobile use due to required size and power. Another important factor is backward compatibility – a mobile 3D display should support both 2D and 3D modes and switch to the correct mode when the respective content is presented.

While selecting the enabling display technology suitable for 3D media handhelds, autostereoscopic displays seem the most adequate choice. These displays create 3D effect requiring no special glasses. Instead, additional optical elements are aligned on the surface of the screen (normally an LCD), to redirect the light rays and ensure that the observer sees different images with each eye [13], [15]. Typically, autostereoscopic displays present multiple views to the observer, each one seen from a particular viewing angle along the horizontal direction. The number of different views comes at the price of reduced spatial resolution and lowered brightness. In the case of small-screen, battery-driven mobile device the trade-off between number of views and spatial resolution is of critical importance. As mobile devices are normally watched by single observer only, two independent views are considered sufficient for satisfactory 3D perception and good compromise with respect to spatial resolution.

A. An overview of portable autostereoscopic displays

Basically, an autostereoscopic display operates by "casting" different images towards each eye of the observer in order to create binocular cues through binocular disparity. This is done by a special optical layer, additionally mounted on the screen surface which controls the light passing through it. The additional layer optically selects different pixels of the conventional LCD or OLED behind it to be included in left or right view. A composite image combining the two views is rendered on the display pixels but only the (sub-) pixels which belong to the correct view are visible to the corresponding eye. There are two common types of optical filters – lenticular sheet and parallax barrier.

Lenticular sheets are composed by small lenses with special shape, which refract the light to different directions [15]. The shapes are formed as cylindrical or spherical in order to enable the proper light redirection. Parallax barrier is essentially a mask with openings and closings which blocks the light in certain directions [16]. In both cases, the intensity of the light rays passing through the filter changes as a function of the angle, as if the light is directionally projected. Each eye sees the display from different angle and thus sees only a fraction of all pixels, precisely those

meant to convey the correct (left or right) view, otherwise combined in the rendered image. The two technologies are illustrated in Figure 8.

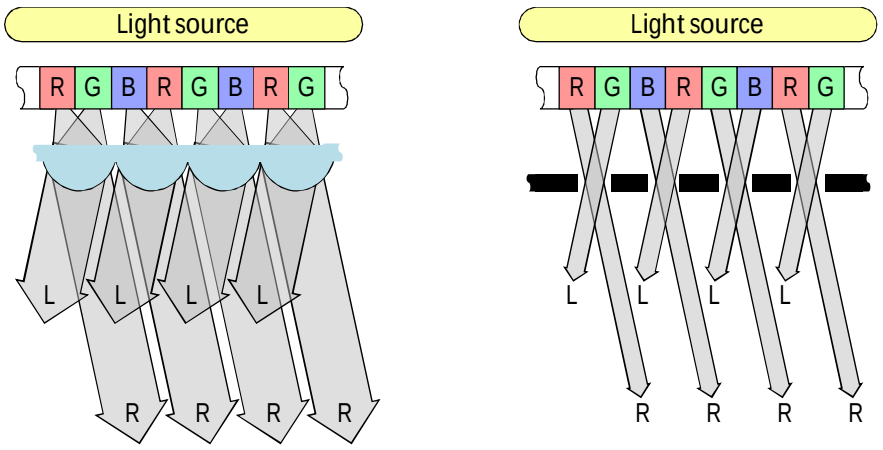


Figure 8. Light redirecting in auto-stereoscopic displays: lenticular sheet (left) and parallax barrier (right).

Both technologies have certain limitations. The viewer should be placed within a restricted area, called a sweet spot, in order to perceive 3D image. Moving outside this proper area, the user might catch the opposite views and experience so-called pseudoscopy. Non-ideal separation between views creates inter-view cross-talk manifested in ghost-like images. This effect occurs especially if the viewer is not in the optimal viewing position. As different sub-pixels are responsible for different-perspective images, the spatial resolution is decreased and the discrete structure of views becomes more visible. Parallax barriers block part of the light and thus decrease the overall brightness. In order to compensate for this limitation, one needs extra bright backlight, which would decrease the battery life if used in a portable device. Nevertheless, auto-stereoscopic displays have been the main candidates for 3D-enabled mobile devices. Amazingly enough, some of the drawbacks of auto-stereoscopic displays in bigger sizes, such as lack of continuous parallax, limited number of different views and inability to serve multiple users, are reduced in their mobile counterpart versions, since typical use scenario assumes single user and no multiple views. In addition, the user can easily adjust the device so to find the correct observation angle.

TFT displays recreate the full color range by emitting light though red, green and blue colored components (sub-pixels). Sub-pixels are usually arranged in repetitive vertical stripes as seen in Figure 9. Since sub-pixels appear displaced in respect to the optical filter, their light is redirected towards different positions. One group will provide the image for the left eye, another – for the right. In order to be shown on a stereoscopic display, the images intended for each eye should be spatially multiplexed. This process is referred to as *interleaving* [1] or *interzigging* [27] and depends on the parameters on the optical filter used. Two topologies are most commonly used. One interleaves on pixel level, where odd and even pixel columns belong to alternative views. The other interleaves on a sub-pixel level – where sub-pixel

columns belong to alternative views. In the second case, different-color components of one pixel belong to different views.

	Pixel 1			Pixel 2			Pixel 3		
Row	R	G	B	R	G	B	R	G	B
1	L	L	L	R	R	R	L	L	L
2	L	L	L	R	R	R	L	L	L
3	L	L	L	R	R	R	L	L	L

	Pixel 1			Pixel 2			Pixel 3		
Row	R	G	B	R	G	B	R	G	B
1	L	R	L	R	L	R	L	R	L
2	L	R	L	R	L	R	L	R	L
3	L	R	L	R	L	R	L	R	L

Figure 9. Interleaving of image for stereoscopic display on pixel level (left) and sub-pixel level (right).

The first display for a mobile phone was announced by Sharp Laboratories of Europe in 2002 [17]. Since then a few vendors announced prototypes of 3D displays, targeted for mobile devices [18], [19], [20]. All of them are two-view, TFT-based autostereoscopic displays. The display produced by Sharp uses electronically switchable reconfigurable parallax barrier, working on sub-pixel basis [17]. The interzigging topology is similar to the one of Figure 9 left. Each view is visible from multiple angles, and the angle of visibility of one view is rather narrow, making the visual quality of the 3D scene quite sensitive to the observation angle.

Another 3D-LCD module based on switchable parallax barrier technology has been produced by Masterimage [20]. It is 4.3" WVGA autostereoscopic display which can operate in 2D or 3D mode. The parallax barrier of the 3D LCD module can be switched between "3D horizontal" and "3D vertical" mode, allowing it to operate in landscape 3D or portrait 3D mode. The barrier operates on pixel level.

From the group of displays based on lenticular lenses, we refer to two prototypes, delivered by Ocuity Ltd. and NEC LCD respectively. The reconfigurable 2D/3D technology by Ocuity Ltd. uses a Polarization Activated Microlens array [19]. The microlens array is made from a birefringent material such that at the surface of the lens there is a refractive index step for only one of the polarizations.

The WVGA 3D LCD module with HDDP (Horizontal Double-Density Pixel) structure as developed by NEC Central Research Laboratories uses NEC's proprietary pixel array for stereoscopic displays [18]. The HDDP structure is composed of horizontally striped RGB colour sub-pixels; each pixel consists of three sub-pixels that are striped horizontally and split in half lengthwise. As a result, horizontal resolution is doubled compared to 3D LCD modules constructed with vertically striped pixels, and 3D images are produced through data for the right eye and data for the left eye being alternately displayed horizontally by pixel. Moreover, 2D images may also be displayed when the same data is presented for adjacent pixels. Since the LCD module can display both 3D and 2D images at the same resolution, it can display a mixture of 2D and 3D images simultaneously on the same screen without causing discomfort to viewers. The pixel arrangement is illustrated in Figure 10.

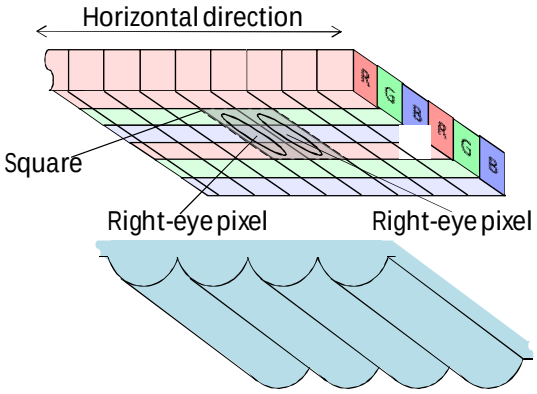


Figure 10. HDDP pixel arrangement.

Last display we overview is produced by 3M. It is based on patterned retardation film, which distributes the light into two perspective views in a sequential manner. The display uses a standard TFT panel operating at 120Hz with special type of backlight. It is composed of two sources of light, a lightguide and 3D film between the LCD and the lightguide. The construction is shown in Figure 11.

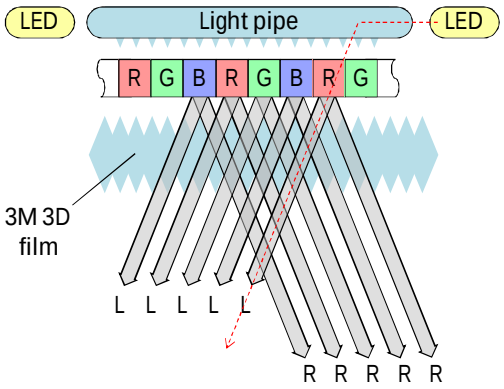


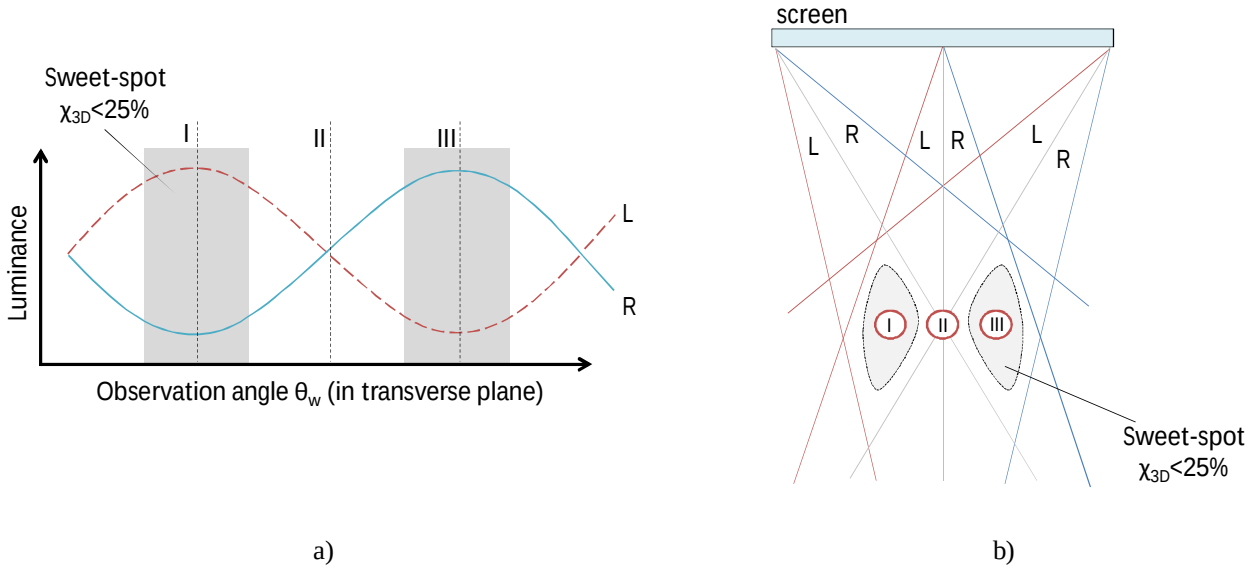
Figure 11. 3D film-based display.

The two backlights are turned on and off in counter phase so that each backlight illuminates one view. The switching is synchronized with LCD, which displays different-perspective images at each backlit switch-on time. The role of the 3D film is to direct the light coming from the activated backlight to the corresponding eye.

B. Optical parameters of portable autostereoscopic displays

Various optical parameters can be used for characterizing the quality of autostereoscopic 3D displays. The set of parameters includes *angular luminance profile* [21], *3D-crosstalk and luminance uniformity* [22], *viewing freedom*, pixel *'blockiness'* and *'stripiness'* [23] as well as angular measurements in Fourier domain [24]. Visual appearance of a 3D scene also depends on external factors, such as observation distance, ambient light and scene content. Therefore, for comparing the visual quality of autostereoscopic displays, one should select the subset of perceptually important optical characteristics.

Crosstalk is perhaps the single most important parameter affecting the 3D quality of autostereosopic displays. For autostereoscopic displays, crosstalk can be calculated as the ratio χ_{3D} of visibility of one view to the visibility to all other views [22]. A number of studies investigated how the level of crosstalk affect the perceptibility of stereoscopic 3D scenes [25][31][40]. According to [25], crosstalk of less than 5% is undistinguishable and crosstalk over 25% severely reduces the perceptual quality. To characterize the influence of cross-talk, one can regard the visibility on the horizontal plane passing through the center of the display, the so-called *transverse plane* [24]. For autostereoscopic 3D displays with no eye tracking, both the luminance of a view and crosstalk between views are functions of the observation angle with respect to that plane, as shown in Figure 12a. For each point on the display surface, there are certain observation angles, where the crosstalk is low enough to allow 3D perception with sufficient quality. The positions, at which one view is seen across the whole display surface have diamond-like shapes on the transverse plane and are called *viewing diamonds* [22][23]. The areas inside the viewing diamonds where the crosstalk is sufficiently low are the *sweet spots* of the views [23]. In Figure 12, areas marked with "I" and "III" are the sweet spots of the left and right views correspondingly. A cross-talk level χ_{3D}<25% can be used to define the sweet spots of the views.



a) b)

Figure 12. a) Angular luminance profile of two-view autostereoscopic display and b) its viewing diamonds.

A set of mobile 3D displays is listed in Table I. The HDDP device uses display with HDDP pixel arrangement [18]. The MI(P) and MI(L) devices use switchable parallax barrier display interleaved on pixel level, operating in portrait and landscape modes correspondingly [20]. The FF [26] and SL [17] devices use switchable parallax barrier interleaved on sub-pixel level. The FinePix camera, designated as FC uses time-sequential 3D-film -based display [26]. As an alternative, measurement results for a row-interleaved, polarization-multiplexed 3D display with glasses (AL) are presented in the last row of the table.

Table 1. Devices with 3D displays used in the measurement tests.

Label	Model	Type	OVD, cm
HDDP	NEC HDDP prototype	3.2" display based on the lenticular HDDP technology by NEC	40
MI(L)	MB403M0117135 – landscape mode	Mobile 3D display with parallax barrier, switchable between landscape and portrait mode	37
MI(P)	MB403M0117135 – portrait mode		
FF	FinePix REAL 3D V1	2D Photo Frame with parallax barrier	46
FC	FinePix REAL 3D W1	Consumer 3D Camera with 3D display	40
SL	Sharp AL3DU	3D laptop with parallax barrier	58
AL	Acer AS5738DG-6165	3D laptop with polarized glasses	60 (nominal)

Due to imperfect display optics the views are never fully separated, and even in the sweet spots some residual crosstalk exists. This effect is referred to as *minimal crosstalk*, and its value determines the visual quality of the display for the optimal viewing angle and distance. The minimal crosstalk for all measured devices is given in Figure 13. The HDDP display has the lowest crosstalk (χ_{3D}=4%), and thus has the best overall quality among the compared displays. On the FinePix 3D display (FC) the crosstalk measurements consistently reached over 30%., manifested in double edges visible at all times, though stereoscopic perception was still possible. Notably, the AL display performs better when watched with its original glasses (χ_{3D}=24%) than when watched with another pair of general purpose polarized glasses (χ_{3D}=29%).

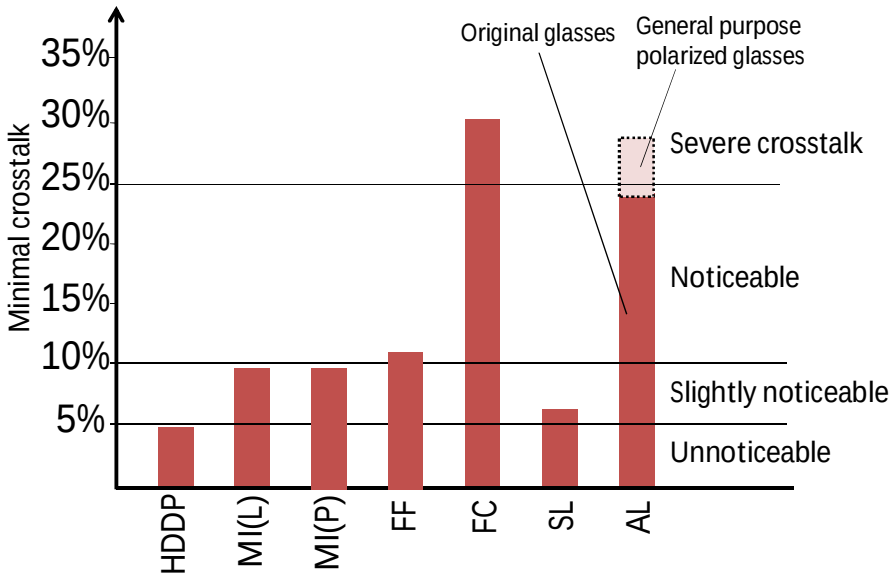


Figure 13. Minimal crosstalk for various mobile 3D displays.

For most autostereoscopic 3D displays the stereoscopic effect can be seen within a limited range of observation distances. The *visibility range* of a 3D display is defined as the range, for which both eyes of the observer would fall into view sweet spot simultaneously. It is limited by the minimum and maximum viewing distances, VD_{min} and VD_{max}

(cf. Figure 14a) while at the OVD the sweet spot has typically the largest width. Usually at this distance the display has the lowest overall crosstalk as well. Since the sweetspots have non-symmetric shape, the interpupilar distance (IP) of the observer affects the VD_{min} and VD_{max} values. Comparative results for IP=65mm and χ_{3D}<25% are given in Figure 14 (see also the measured OVD values in Table I). Since the minimal crosstalk of FC display is always over 30%, from herein it is represented with dashed line, for distances where 30%<χ_{3D}<50%. The AL display does not have neither optimal nor maximal viewing distance in terms of crosstalk. For that display, the OVD is the nominal observation distance as suggested in the display manual.

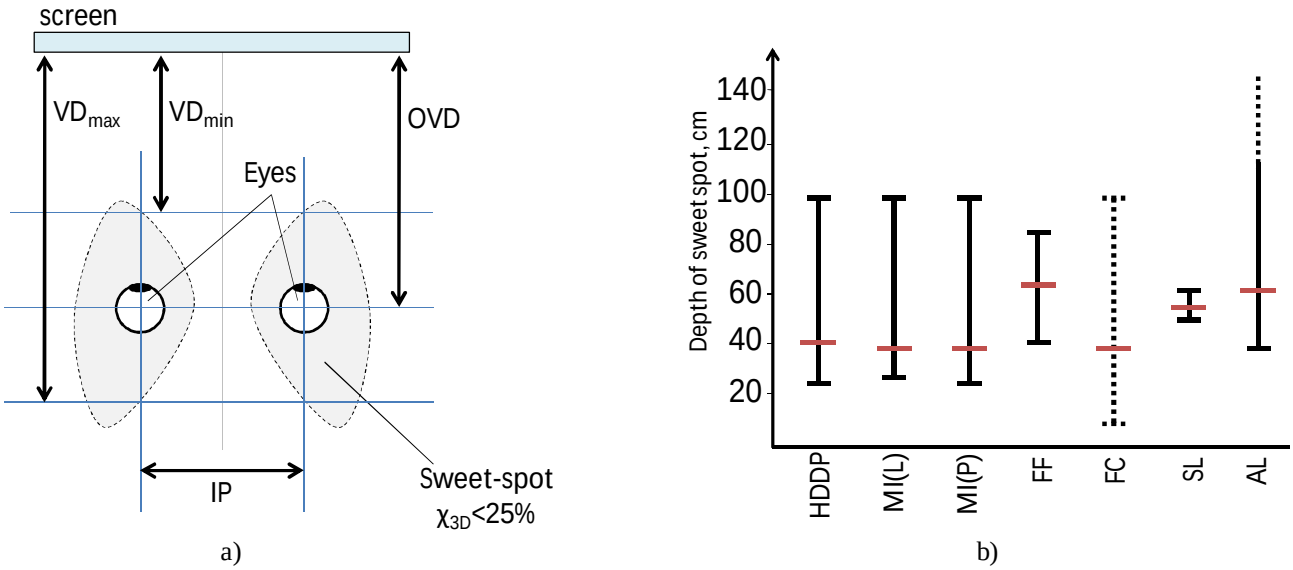
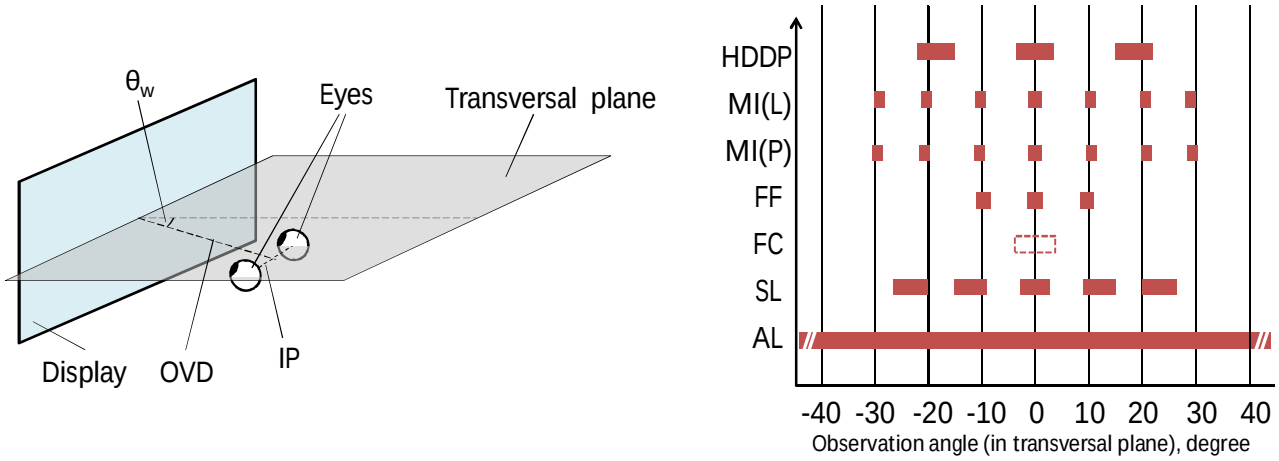


Figure 14. a) Definition of OVD, VDmin and VDmax values; b) measured values for various 3D displays.

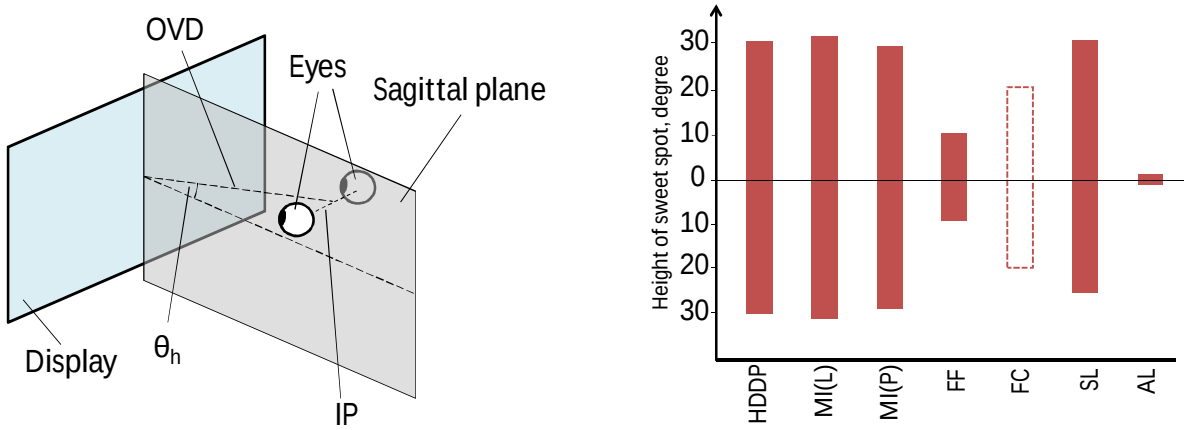
We define the *width of sweet spot* as all angles on the transversal plane, where each eye of the observer perceives the correct view (i.e. not reverse stereo) with crosstalk χ_{3D}<25%. The lateral sweet spot width can be measured in distances, as in [22] and [23]. However, assuming that the observer is always at the optimal distance from the center of the display, the ranges can be measured also in angles, as illustrated in Figure 15a. This is done as it is more likely that the user of a mobile display is holding it at a constant distance, and is turning it in order to get the best view. Typical results for IP=65mm are given in Figure 15b. Among all autostereoscopic displays tested, HDDP has the widest sweet spots, which makes it the easiest for the user to find a correct observation angle. On contrary, the MI display has narrow sweet spots and users must hold it at a precise angle to be able to perceive stereoscopic effect. The AL display used with glasses delivers continuous 3D effect over a wide range of observation angles.



a) b)

Figure 15. a) Measurement of sweet spot width and b) sweet spot widths for various mobile 3D displays.

The *sweet spot height* is measured as the range of observation angles in the plane passing through the center of the display (also known as sagittal plane), where observers' eyes perceive correct stereo with χ_{3D}<25%. The user is assumed to be at the display's OVD, as shown in Figure 16a. The measurement results for IP=65mm are given in Figure 16b. Most autosteoscopic displays have vertical observation range of -30 to 30 degrees. Interestingly enough, the AL display is very sensitive to the vertical angle, and has a sweet spot height of -2 to 2 degrees. In fact, this is the limiting factor defining the minimum observation distance for that display.



a) b)

Figure 16. a) Measurement of sweet spot height and b) sweet spot heights for various mobile 3D displays.

In contrast to 2D displays, where the user is free to choose the observation distance, autostereoscopic 3D displays deliver best results when observed at their OVDs. Since OVD varies from display to display, it is more suitable to compare angle-of-view (AOV) and angular resolution, rather than the absolute size and resolution of such displays. The area, which each display occupies in the visual field, when observed from its optimal observation distance is given in Figure 17a. Next to each display is given its OVD. The angular size of all displays, observed at their OVD is given in Figure 17b. For MI, FF and SL displays, both results for 2D and 3D modes are given as the resolutions are different. For comparison, the angular resolutions for the displays of two popular handhelds, Nokia N900 and Apple iPhone4, at

40cm observation distance are given. The theoretical angular resolution of the human retina (50CPD) is calculated for perfect 20/20 eyesight. Figure 17 is instructive about the fact that 2D and 3D displays have comparable AOV but different angular resolution. Especially the horizontal angular resolution of mobile 3D displays is much lower than the one of a typical mobile 2D display.

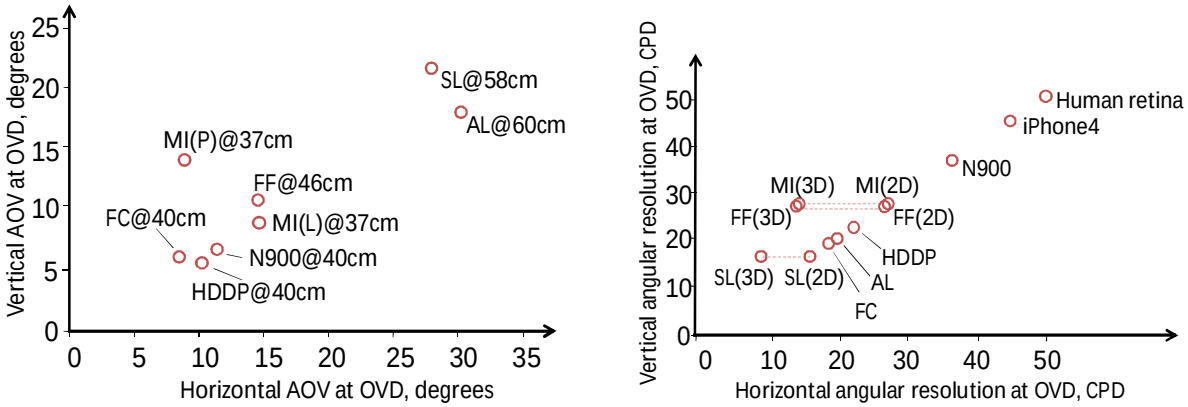


Figure 17. Angular size and angular resolution of various mobile 3D displays: a) angular size observed from OVD, in degrees; b) angular resolution observed from OVD, in cycles per degree. Note: N900 and iPhone4 are 2D displays given for comparison, as they appear at 40cm observation distance.

IV. USER EXPERIENCE OF 3D MEDIA FOR MOBILES

User experience seems to be the key factor for the adoption of the mobile 3D media technology, as having a perceptually acceptable and high-quality 3D scene on a small display is a challenging task. According the holistic user-centered research framework, as formulated in Section II.B, research efforts have focused on optimizing the technology components, such as content creation and coding techniques, delivery channels, portable 3D displays and media-rich embedded platforms to deliver the best possible visual output. In this section, the 3D media user experience is addressed methodologically by an interdisciplinary approach having three-fold goals. First, the artifacts, which arise in various usage scenarios involving stereoscopic content, are analyzed and categorized so to put them against the peculiarities of the human visual system and the way users perceive depth. Then, critical parts of the system, such as coding and transmission approaches, are studied for their performance both through objective comparisons and subjective tests so to reveal the levels of acceptance and satisfaction of the new content and services. Eventually, 3D graphical user interfaces complement the experience of 3D media content.

A. 3D-specific artifacts

Stereoscopic artifacts can be described with respect to the stage in the 3D media delivery chain, as exemplified in Figure 5 and how they affect different “layers” of human 3D vision. In this way, artifacts can be clustered in a multidimensional space according their source and *structure, color, motion and binocular* “layers” of HVS, interpreting

them. These layers roughly represent the visual pathways as they appeared during the successive stages of evolution. The *structure* layer denotes the spatial and colorless vision. It is assumed that during the evolution human vision adapted for assessing the "structure" (contours and texture) of images [35], and some artifacts manifest themselves as affecting image structure. *Color* and *motion* layers represent the color and motion vision, correspondingly. The *binocular* layer denotes artifacts meaningful only when perceived in a stereo-pair, and not by a single eye (e.g. vertical disparity). The result of multidimensional clustering is well illustrated by a circular diagram in polar coordinates given in Figure 18, [39]. Such a wide nomenclature of clustered artifacts helps in identifying the stages at which they should be properly tackled. While some of the artifacts are less important in mobile context, some are quite typical and influential for the acceptance of the technology.

Artifacts caused at creation/capture stage

The most common and annoying artifact introduced in the process of capture or rendering a stereoscopic image is *unnatural disparity* between the images in the stereo-pair. Special care should be taken when positioning cameras or when selecting rendering parameters and rectification is a standard pre-processing stage. However, often a perfectly rectified stereoscopic image needs to be visualized at different size than the originally captured one. Changing the size or resolution of stereoscopic pair can also introduce unnatural disparity. When resizing a stereoscopic pair, the relative disparity is scaled proportionally to the image size. However, as the interocular distance remains the same, observing a closely-positioned mobile 3D display would require different relative disparity range compared to when observing large 3D display placed further away. The effect is illustrated in Figure 19. Even if the mobile and large 3D displays have the same visual size, stereoscopic images on them have different disparity.

Two-channel stereo video, and video plus dense depth are the likely contenders for 3D video representation for mobiles [1]. If the representation format is different from the one the scene has been originally captured, converting between the formats is a source of artifacts. A typical example is the occlusions areas in depth-from-stereo type of conversion.

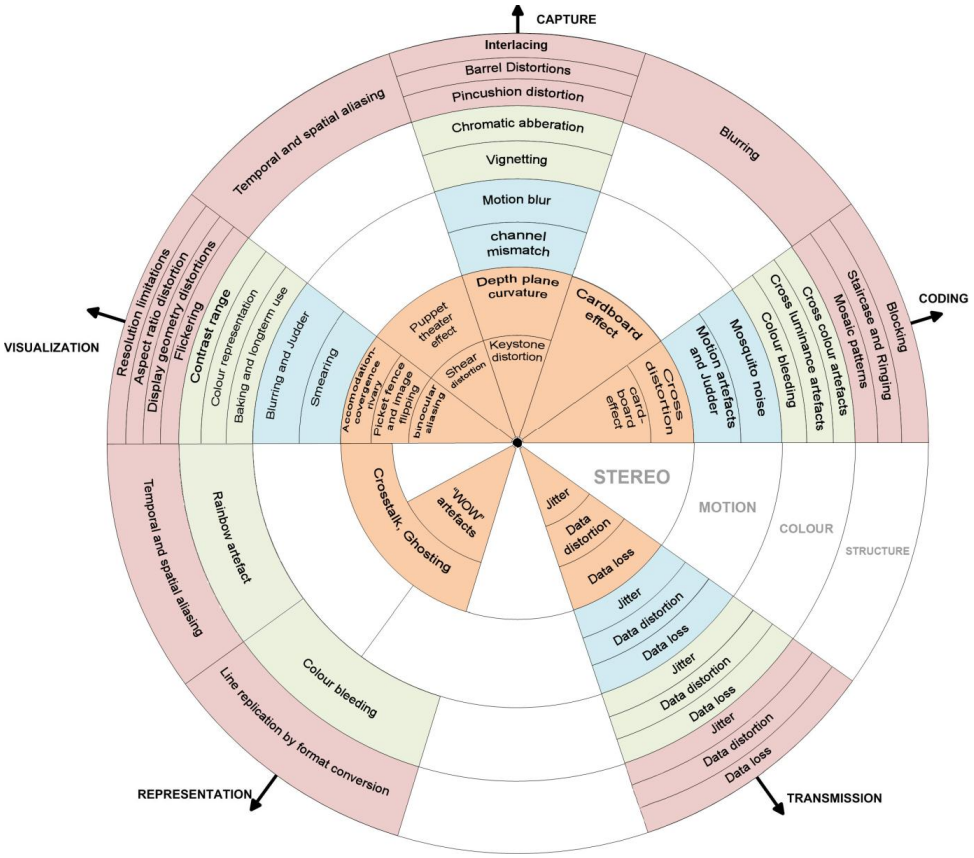


Figure 18. Artifacts, caused by various stages of content delivery and affecting various "layers" of human depth perception.

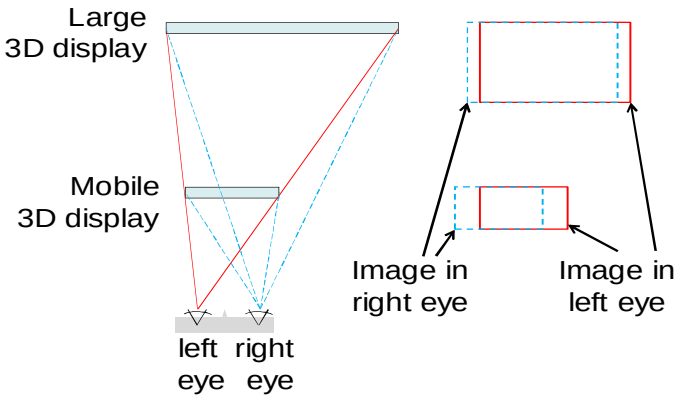


Figure 19. Change of relative disparity while rescaling stereoscopic image pair.

Coding artifacts

Various coding schemes, utilize temporal, spatial or inter-channel similarities of a 3D video [2]. Algorithms originally designed for single-channel video, might be improperly applied for stereo-video, and important binocular depth-cues might be lost in the process. Using block-based DCT compression is a source of *blocking artifacts*, which are thoroughly studied for 2D video, but their effect on stereoscopic quality is yet to be determined. Some authors propose that blocking might be considered as several, visually separate artifacts – *block-edge discontinuities*, *color bleeding*, *blur* and *staircase artifacts* [35], [36]. Each of these artifacts introduces different amount of impairments to object edges

and texture. The human brain has the ability to perceive single image by combining the images from left and right eye (so-called cyclopean image) [33]. As result, the same level of DCT quantization might result in different perceptual quality, based on the depth cues present in a stereo image. In Figure 20, both channels of a stereo-pair are compressed with the same quality factor. When an object appears on the same place in both frames, it is equally affected by blocking in each frame, and the perceived cyclopean image is similar to the one shown in Figure 20a. When the object has different horizontal position in each frame, the blocking artifacts will affect differently the object in each frame, which results in a cyclopean image similar to the one in Figure 20b.

Figure 20. The impact of blocking on stereopairs with different disparity: a) q=15, disparity=0; b) q=15, disparity=4; c) zoomed detail of a); d) zoomed detail of b).

Transmission artifacts

In the case of digital wireless transmission a common problem is packet losses. Related artifacts are sparse, and highly variant in terms of occurrence, duration and intensity. At very low bit rates they may be masked by compression impairments. The presence of artifacts depends very much on the coding algorithms used and how the decoder copes with the channel errors. In DVB-H transmission most common are burst errors, which results in *packet losses* distributed in tight groups [66]. In MPEG-4 based encoders packet losses might result in *propagating* or *non-propagating* errors, depending on where the error occurs in respect to key frames, and the ratio between key- and predicted frames. Error patterns of wireless channels can be obtained with field measurements, and then used for simulation of channel losses [66], [67]. In multiview video encoding, where one channel is predicted from the other, usually error burst is long enough to affect both channels [68]. In that case, packet loss artifacts appear on the same absolute position in both images even though the appearance in one channel is mitigated due to the prediction. Figure 21 illustrates the effect for the case of TU6 channel with channel SNR=18 dB [68]. In the format V+D using a separate depth channel, usually depth is encoded in much lower bitrate than the video. In that case burst errors affect mainly the video channel, and the relative perceptual contribution of depth map degradation alone is very small.

Figure 21. Packet loss artifacts affecting multi-view encoded stereoscopic video [68].

One common artifact introduced during receiving and decoding of 3D video, is *temporal mismatch*, where one channel gets delayed in respect to the other. It might be caused by insufficient memory or CPU, or error concealment in one channel. The outcome is that the image from one channel do not appear with simultaneously taken image from the other channel, but with an image which is taken a few frames later. Even temporal mismatch of as low as two frames can result in a stereoscopically inadequate image pair. For comparison, two images are shown in Figure 22 - the left is done by superimposing the frame 112 from left and right channels of a movie; the right is done by superimposing a frame 112 from the left channel and frame 115 from the right channel of the same movie.

Figure 22. Temporal mismatch in stereo-video. Left: superimposed images of temporally-synchronized stereopair, right: superimposed images of stereopair with 3 frames temporal mismatch.

Visualization and display artifacts

Even a perfectly captured, transmitted and received stereoscopic pair can exhibit artifacts due to various technical limitations of the autostereoscopic display in use [69], [70], [71]. The most pronounced artifact in autostereoscopic displays is cross-talk, caused by imperfect separation of the “left” and “right” images and is perceived as *ghosting artifacts* [27]. Two factors affect the amount of crosstalk introduced by the display – position of the observer and quality of the optical filter in front of the LCD, as discussed in Section III.B. Due to the size of the sub-pixels, there is a range of observation positions, from where some sub-pixels appear partially covered by the parallax barrier, or are partially in the focal field of the corresponding lenticular lens. This creates certain *optimal observation spots* in the

centers of the sweet spots, where the two views are optimally separated (the areas marked with “I” and “III” in Figure 12b), and transitional zone (marked with “II”) where a mixture of the two is seen. However, even in the optimal observation spot one of the views is not fully suppressed – for example part of the light might “leak” through the parallax barrier as shown in Figure 23a and create the minimal crosstalk effect discussed in Section III.B.

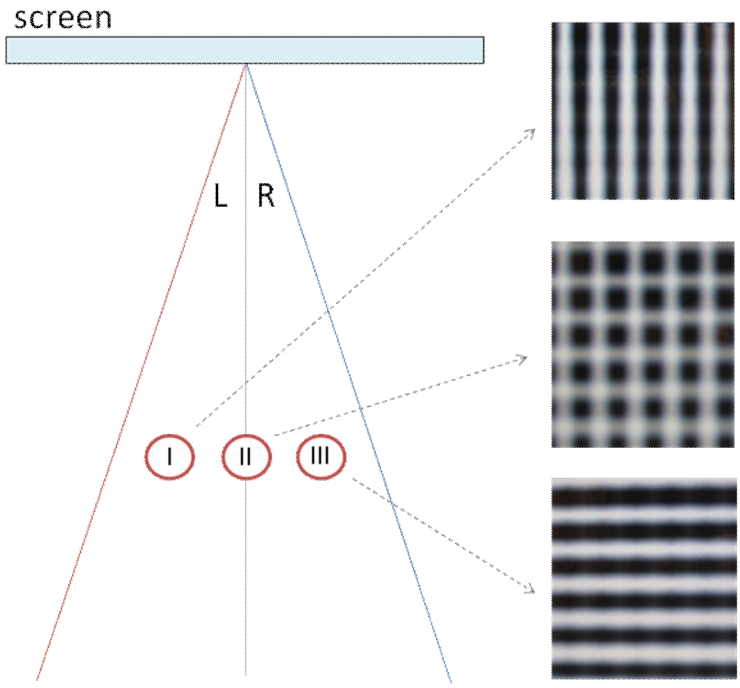


Figure 23. Effect of crosstalk in portable 3D displays; from left to right - photographs taken of a 3D display from positions I, II and III.

The effect is well illustrated by a special test stereoscopic pair, where the “left” image contains vertical bars, and the “right” image contains horizontal bars. This stereo pair has been visualized on a parallax-barrier based 3D display, and photographed from observation angles as marked with “I”, “II” and “III” in Figure 12a. The resulting photos are shown in Figure 23c, d and e. Both position-dependent and minimal crosstalk effects can be seen. By knowing the observation position and the amount of crosstalk introduced by the display, the effect of crosstalk can be mitigated by pre-compensation [146].

There are darker gaps between sub-pixels of an autostereoscopic display. They are more visible from certain angles than from others. When an observer moves laterally in front of the screen, he perceives this as luminance changes creating brighter and darker vertical stripes over the image. Such effect is known as *banding artifacts* or *picket fence effect* and is illustrated in Figure 24. The effect can be reduced by introducing a slant of the optical filter in respect to the pixels on the screen [15]. Tracking of the user position in respect to the screen also can help reducing these artifacts.

Figure 24. Banding / picked fence artifacts.

Parallax-barrier and lenticular based 3D displays with vertical lenses arrangement have horizontal resolution twice lower than vertical one as only half of the sub-pixels of a row form one view. This arrangement requires spatial sub-sampling of each view, before both views are multiplexed, thus risking introducing aliasing artifacts. In 3D displays, aliasing might cause false color or Moiré artifacts (illustrated in Figure 25) depending on the properties of optical filter used. Properly designed pre-filters should be used, in order to avoid aliasing artifacts.

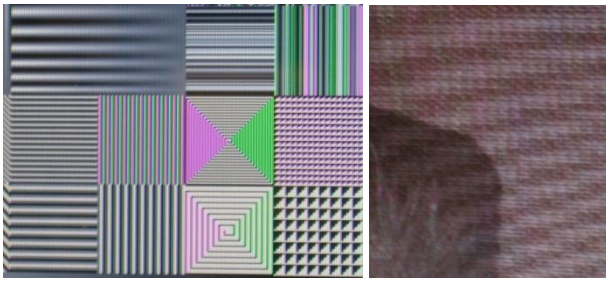


Figure 25. Aliasing in autostereoscopic displays: left: false color; right: Moiré artifacts.

Autostereoscopic displays which use parallax barrier usually have a number of interleaved “left channel” and “right channel” visibility zones, as shown in Figure 26. Such display can be used by multiple observers looking at the screen at different angles, for example positions marked with “1” and “2” in the figure. However, an observer in position “3” will perceive *pseudoscopic* (also known as *reversed stereo*) image. For one observer, this can be avoided by using face tracking and algorithm which swaps the “left” and “right” images on the display appropriately to accommodate to the observers viewing angle.

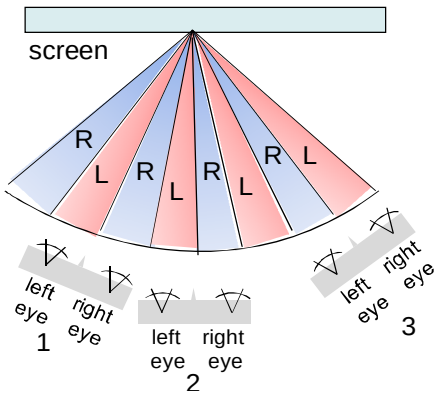


Figure 26. True stereoscopic (1 and 2) and pseudoscopic (3) observation positions.

B. Optimized delivery channel

Evaluation of coding methods

The methods for 3D video coding described in Section II.C contain a multitude of parameters which vary their performance in different scenarios. As all methods are based on H.264 AVC, the profiles of the latter (i.e. Baseline, Main, Extended, and High Profiles), its picture type (I, P and B) and entropy coding methods (CABAC or CAVLC) determine the varying settings to be tested for mobile use [83].

In [84], candidate stereoscopic encoding schemes for mobile devices have been investigated for both encoding and decoding performance. Rate-distortion curves have been used to assess the coding efficiency and decoding speed tests have been performed to quantify the decoder complexity. It has been concluded that, depending on the processing power and memory of the mobile device the following two schemes can be favored: H.264/AVC MVC extension with simplified referencing structure and H.264/AVC monoscopic codec with IPP+CABAC settings over interleaved stereoscopic content.

In [85], H.264/AVC simulcast, H.264 Stereo SEI message, H.264/MVC, MPEG-C Part 3 using H.264 for both video and depth and H.264 auxiliary picture syntax for video plus depth have been compared for their performance in mobile setting. A set of test videos with varying type of content and complexity has been used. The material has been coded at different bitrates using optimum settings for each of the above mentioned encoders. The quality has been evaluated by means of PSNR over bitrate. The results show that the overall RD-performance of MVC is better than simulcast coding. It has also been shown that the overall RD-performance of video plus depth is better than stereo video with simulcast coding.

The selection of an optimum coding method has recently been addressed in two publications by Strohmeier and Tech [121], [124] based on the results from subjective tests. Four different coding methods that had been adapted for 3D

mobile television and video were evaluated. H.264/AVC Simulcast [133], H.264/MVC [134], and Mixed Resolution Stereo Coding (MRSC) [127], [128] using H.264/AVC were chosen as coding methods for a Video + Video approach. Video plus Depth Coding using MPEG-C Part 3 [135] and H.264/AVC as a Video + Depth approach completed the coding methods under assessment. The depth maps of the test sequences were obtained using the hybrid-recursive-matching algorithm, described in [147]. The virtual views were rendered following the approach described in [148]. To further decrease the coding complexity with regard to limited calculation power of current mobile devices, the baseline profile was used for encoding. This includes a simplified coding structure of IPPP and the use of CAVLC. Six different contents were encoded at two different quality levels. To determine the different quality levels, the quantization parameters (QP) of the encoder for Simulcast Coding were set to 30 for the high quality and 37 for the low quality. From these sequences, target bit rates for the other methods were derived and used in the test set creation, respectively. Table 2 presents the target bitrates for different quality levels and contents.

Table 2 Target bitrates for different quality levels and test contents

Profile	Quality	Bullinger	Butterfly	Car	Horse	Mountain	Soccer2
Baseline	Low	74	143	130	160	104	100
	High	160	318	378	450	367	452
High	low	46	94	112	104	78	134
	High	99	212	323	284	208	381

The test items were evaluated by 47 test participants. The evaluation followed the Absolute Category Rating (ACR) [115] and test participants evaluated acceptance of (yes/no) and satisfaction with (11-point-scale) perceived overall quality. The test items were presented using a NEC HDDP 3.5” mobile display [136] with a resolution of 428 x 240 px.

All coding methods under test provided a highly acceptable quality at the high quality level of 80% and higher. At low quality level, MVC and V+D still got an acceptance score of 60% and higher. Strohmeier and Tech [121] showed in their study that MVC and the Video + Depth provide the best overall quality satisfaction for both quality levels (see Figure 27). These coding methods significantly outperform MRSC and Simulcast. With respect to the different test contents the results show that coding methods show content-dependent performance. Video + Depth gets the highest overall satisfaction scores for Car, Mountain, and Soccer2. MVC outperforms all other coding methods for content Butterfly.

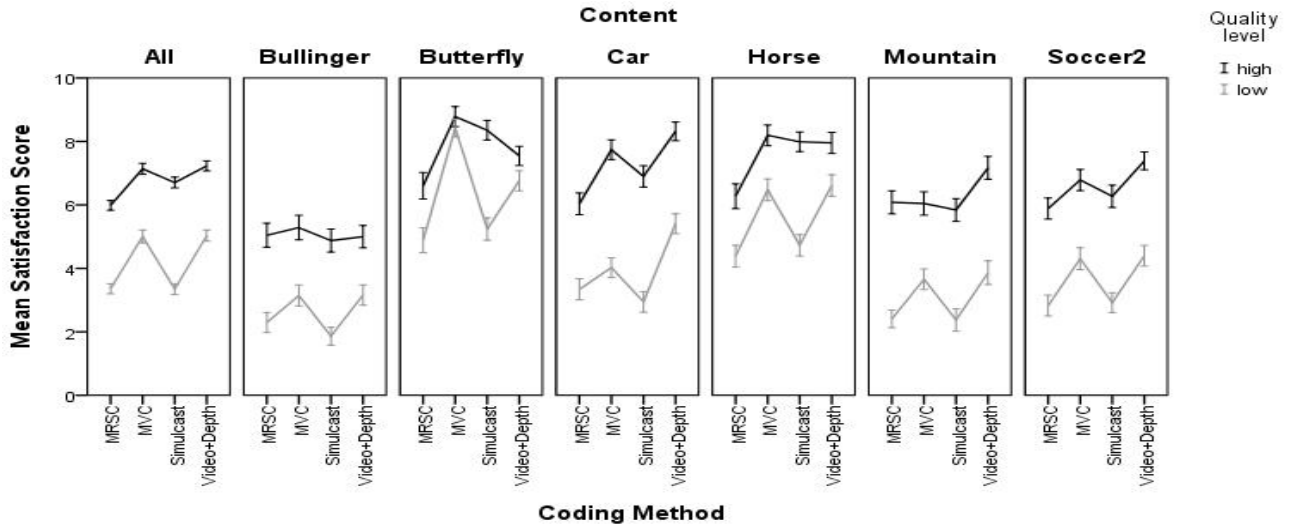


Figure 27. Mean satisfaction scores for different coding methods at baseline profile averaged over contents (All) and content-by-content given at high and low quality levels. Error bars show 95% confidence interval (CI) of mean.

The results of this study were extended in a follow-up study by Strohmeier and Tech [124]. While the first study was limited to the use of low coding complexity, the second study used the complex high profile which enables hierarchical B-frames and CABAC. The other parameters, quality levels, test contents, and device were the same so that the follow-up study [124] allowed a direct comparison of the results of baseline and high profile. 40 participants evaluated the test set of high profile.

The results of the overall quality evaluation for the high profile sequences confirmed the findings of the baseline sequences. As seen in Figure 28, the test items at high quality level got an overall quality acceptance score of at least 75%. For low quality level, MVC and Video + Depth reach an acceptance level of 55% and more. Like in the baseline case, MVC and Video + Depth also outperform the other coding methods in terms of satisfaction with overall quality. The content-dependent results for the provided overall quality for all coding methods were shown in the results as well.

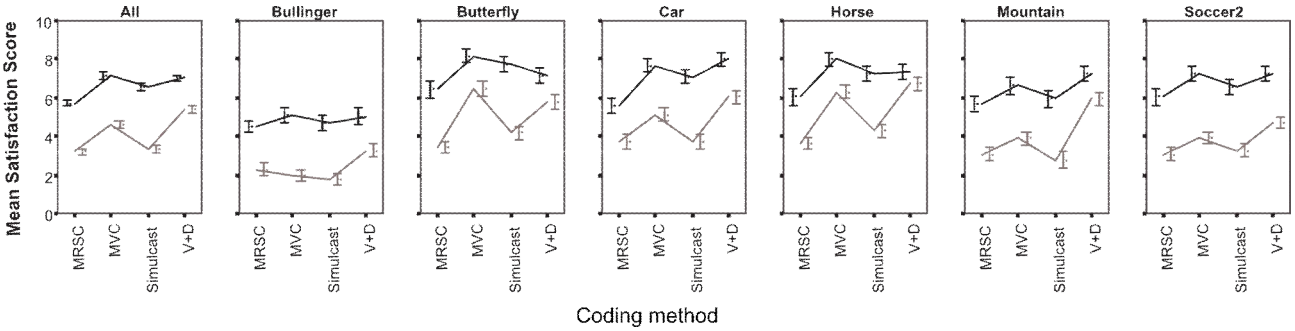


Figure 28. Mean satisfaction scores for different coding methods at high profile averaged over contents (All) and content-by-content given at high and low quality levels. Error bars show 95% CI of mean.

Finally, the results of both studies allowed to directly comparing the performance of baseline and high profiles (see Figure 29). Although the results show small differences for baseline and high codec profiles for some settings, the overall view on the results shows no differences among the two profiles. However, significantly lower bit rates can be realized for the high profile due to more efficient, though more complex, coding structures. Altogether, Strohmeier and Tech [124] showed that the use of high coding profile, i.e. hierarchical b-frames and CABAC, can provide the same experienced quality than baseline profile using lower bit rates. This can result in advantages for the transmission of these sequences in terms of better error resilience [137].

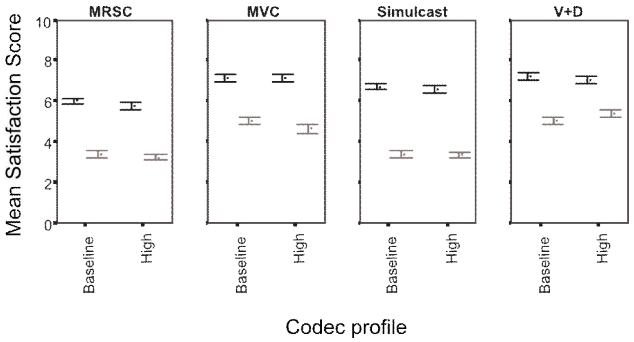


Figure 29. Comparison of the mean satisfaction scores for coding methods used in two studies [121], [124] for baseline and high profile. Error bars show 95% CI of mean.

Evaluation of transmission approaches

In order to illustrate the effects of channel characteristics on the received video quality, a typical 3D broadcasting system is simulated as shown in Figure 30 [96]. In this study DVB-H is used as the underlying transmission channel. DVB-H is the extension of DVB Project for the mobile reception of digital terrestrial TV. It is based on the existing DVB-T physical layer with introduction of two new elements for mobility: MPE-FEC and time slicing. Time slicing enables the transmission of data in bursts rather than a continuous transmission; explicitly signaling the arrival time of the next burst in it so that the receiver can turn on between and wake up before the next burst arrives. By this way the power consumption of the receiver is reduced. Multi-Protocol Encapsulation is used for the carriage of IP datagrams in MPEG2-TS. IP packets are encapsulated to MPE sections each of which consisting of a header, the IP datagram as a payload, and a 32-bit cyclic redundancy check (CRC) for the verification of payload integrity. On the level of the MPE, an additional stage of forward error correction (FEC) can also be added.This technique is called MPE-FEC and improves the C/N and Doppler performance in mobile channels. To compute MPE-FEC, IP packets are filled into an N x 191 matrix where each square of the matrix has one byte of information and N denotes the number of rows in the matrix. The standard defines the value of N to be one of 256, 512, 768 or 1024. The datagrams are filled into the matrix column-wise. Error correction codes (RS codes) are computed for each row and concatenated such that the final size of the matrix is of size Nx255. To adjust the effective MPE-FEC code rate, padding or puncturing can be used. Padding refers to filling the application data table partially with the data and the rest with zero whereas puncturing refers to discarding some of the rightmost columns of the RS-data table.

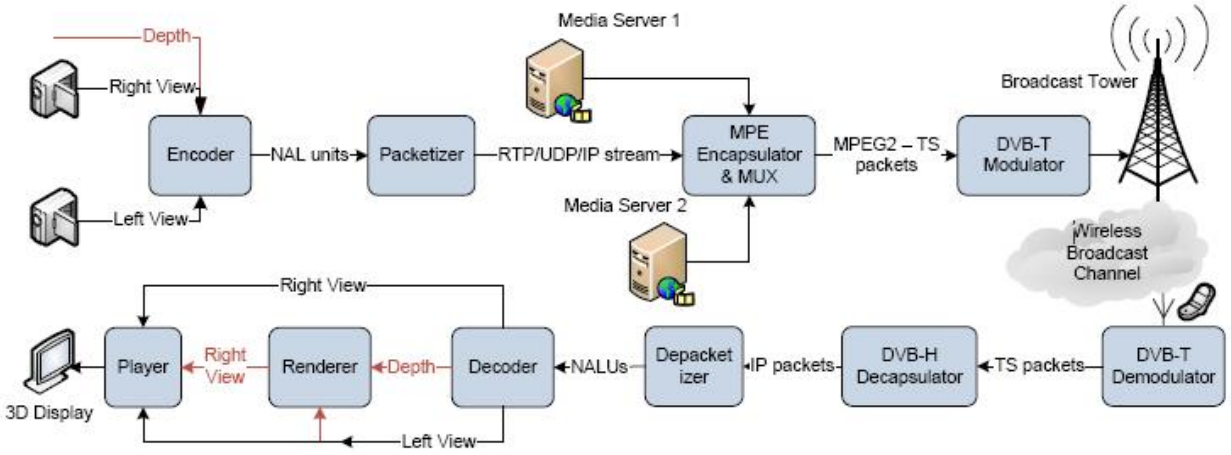


Figure 30. Block diagram of 3D broadcasting system over DVB-H.

In the simulated system, 3D video content is first compressed with a 3D video encoder, operating in one of the modes: MVC, V+D or simulcast. Resulting Network Abstraction Layer (NAL) units (NALU) are fed to the stereo video streamer. The packetizer encapsulates the NAL units into Real Time Transport Protocol (RTP) [95] mono-compatible only, User Datagram Protocol (UDP) and finally Internet Protocol (IP) datagram for each view separately. The resulting IP datagram are encapsulated in the DVB-H link layer where the Multi Protocol Encapsulation Forward Error Correction (MPE-FEC) and time slicing occurs [97]. Through the MPE-FEC mechanism, IP datagrams are protected by adding additional bytes for a variable-length Reed-Salomon (RS) coding. MPE-FER rate refers to the ration between application and total data. Time slicing allows sending the packets into time slices (bursts) for better power consumption at the receiver site. Different views are assigned different PIDs and encapsulated as different elementary streams. Therefore, they are transmitted in different time slices or bursts. The link layer output MPEG-2 Transport Stream (TS) packets are passed to the physical layer where the transmission signal is generated with a DVB-T modulator. After the transmission over a wireless channel, the receiver receives distorted signal and possibly erroneous TS packets are generated by the DVB-T modulator. The received stream is decoded using the section erasure method, i.e. the MPE-FEC frame is filled with contents of the error-free MPE and MPE-FEC sections and the empty bytes in the frame are marked as erasures, RS decoding is performed to reconstruct the lost data, and finally, the received and correctly reconstructed IP datagram are passed to the video client. IP datagram are handled in the depacketizer and resulting NAL units are decoded with the stereo video decoder to generate right and left views. Finally, these views are combined with a special interleaving pattern to be displayed in stereo 3D on an auto-stereoscopic display.

Within the Mobile3DTV project, extensive sets of tests have been performed in order to find an effective compromise between compression efficiency, FEC-code rates and robustness with respect to typical channel conditions [99].

Simulations have been carried out involving 3D video content with different characteristics as described in Table 3 and coded as simulcast, V+D and MVC simplified structure. For all the tests, JMVC 5.05 (in monoscopic mode for simulcast) is used with a GOP size of 8. The quantization parameters (QP) of the encoder are adjusted such that the total bitrate does not exceed 300kbs. For each coding structure, equal error protection (EEP) and unequal error protection (UEP) is applied at the link layer. For EEP, the left and right or video and depth bursts are protected with the same FEC-rate. On the other hand, UEP requires the video bit streams to be partitioned into different segments with different priorities. Segments are then protected with unequal amount of FEC data. For partitioning the video bit streams, there are several methods such as data partitioning and spatial-temporal-quality layering [100]. In the referred study, a partitioning based on the views only is performed, i.e. left/ right views in different segments or left/depth data in different segments. More complex partitioning can also be applied to the stereo data. Once segmented, several unequal protection schemes (UEP) are derived where the channel coding ratio among the streams is determined according to the priority level of the streams.

Table 3 Spatial and temporal characteristics of test contents used in transmission tests

Content	Characteristics	Width	Height	Fps
HeidelbergAlleys	Low camera motion, Low Motion, High Detail	432	240	12.5
KnightsQuest	Computer – Generated,	432	240	12.5
RhineValleyMoving	High Camera and Object Motion, Low Detail	432	240	12.5
RollerBlade	Stationary camera, High Object Motion	320	240	15

In the transmission experiments conducted, a constant typical FEC rate (3/4) is chosen to protect the left and right bursts in the EEP mode since applying an MPE-FEC code rate below R=3/4 at a medium frame size is not recommended without further measures [102]. Then several unequal protection schemes are derived using this EEP structure. Using the FEC rate chosen, 1/4, 2/4, 3/4 and 1 % of the RS columns of right burst (right view or depth) are transferred to the left burst (left view) respectively corresponding to the UEP1, UEP2, UEP3 and UEP4.

For simulating the physical transmission channel, a MATLAB/Simulink tool that models the DVB-T/H modulation and demodulation processes and the physical transmission channel has been used [101]. The channel is modeled as multi-path Rayleigh fading channel with Additive White Gaussian Noise. A mobile use case with Cost 207 radio channel model TU6, having maximum Doppler frequency of 24 Hz is used to obtain the channel specific error patterns. These patterns are then used for modeling the TS packet loss due to channel conditions.

In all the simulations, PSNR values have been used as the distortion metric. First, mean squared errors (MSE) are calculated individually for the left and right channel. They are used to calculate the PSNR for the left and right channel and the average of the two MSEs is used to calculate the average PSNR. At this point we would like to mention that perceptually-driven objective quality metric for stereo images would be more appropriate for comparison than PSNR. There has been an active research toward developing such metrics however they are still deficient in delivering simple, interpretable and reliable results for the mobile case of interest [149].

In case of V+D sequences, since even for the lossless case there is an existing distortion (for PSNR metric) due to imperfections during depth estimation and rendering, original right view is not taken as the reference sequence. Instead , the distortion of the V+D transmissions are given as the PSNR of the received left sequence using original left view as reference; and the PSNR of the right sequence rendered from the received left and depth views using the right view rendered from original left and original depth.

Figure 31and Figure 32 show the PSNR results for different coding and protection methods and for the *RollerBlade* and *KnightsQuest* videos. The results show that MVC performs better than simulcast because of the compression efficiency (bitrate of MVC coded video is chosen to be equal to that of simulcast coded video). UEP in general results in rather marginal improvement over EEP especially under low channel SNR. Also it has been shown that the results depend heavily on the content. If the depth map is accurate as seen in the *RollerBlade* video, V+D representation outperforms the other methods. If the depth map is not accurate, MVC outperforms V+D representation for high SNR cases, however due to the compression efficiency of V+D representation it yields better results for low SNR. On the other hand, at the receiver side, view synthesis has to be performed after decoding to generate the second view of the stereo pair which is rather challenging for mobile devices to achieve in real time.

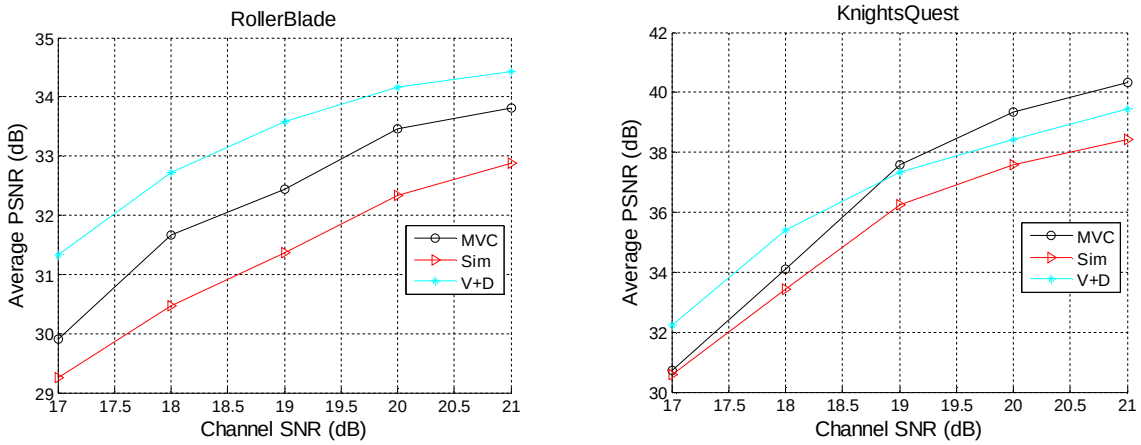


Figure 31 Average PSNR results for coding method comparison in EEP mode.

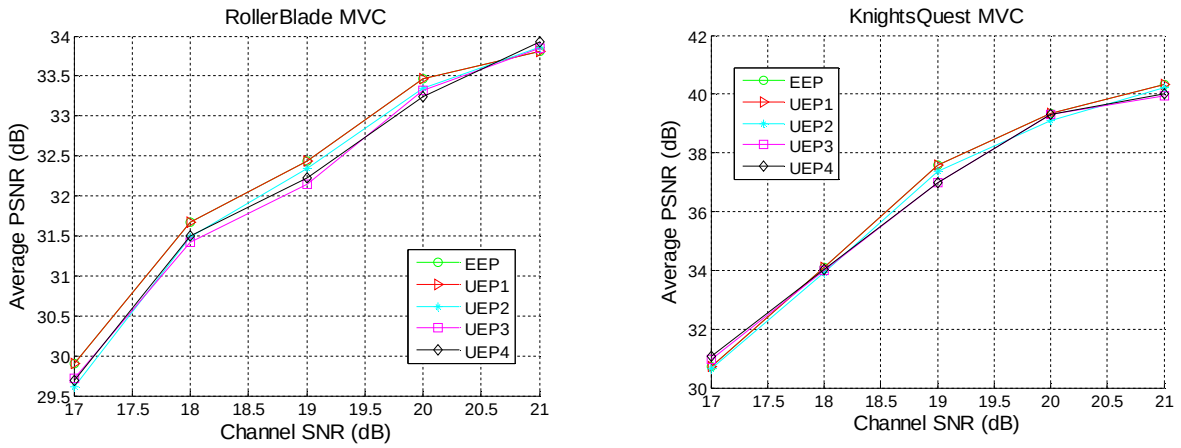


Figure 32. Average PSNR results for protection method comparison in MVC mode.

Subjectively, transmission parameters for mobile 3D media have been evaluated by Strohmeier et al. [122] under the constraint of the studies on coding methods for mobile 3D video by Strohmeier and Tech [121], [124]. This large-scale study has targeted channel transmission parameters taking into account different error resilience methods at the link layer (equal and unequal MPE-FEC) of the DVB-H channel. Regarding the transmission channel, equal (EEP) and unequal (UEP) error protections have been assessed at two different error rates of 10% and 20% corresponding to low and high channel SNRs. According to the results of the coding methods evaluation study, only MVC and the Video + Depth approach were evaluated for all parameters. Four different contents chosen to match the user requirements of mobile 3DTV [116], [117] have been used. 77 test participants took part in the subjective quality evaluation. Absolute Category Rating was chosen as test method and test participants again evaluated acceptance of (yes/no) and satisfaction with (11-point-scale) perceived overall quality.

The results of the study (see Figure 33) [122] confirm the findings of Strohmeier and Tech [121], [124]. At low error rates, the acceptance rate for all test items was at least 60%. The results of the acceptance rate are promising that the current state-of-the-art in mobile 3D media encoding and transmission can already reach a good quality acceptance at the end user. Genuinely 3D coding methods, MVC and Video+Depth, have outperformed Simulcast at all settings. While for low error rates, MVC and Simulcast provided the same quality, MVC has been rated better at higher error rates. Regarding the transmission parameters for low error rates, the results show that MVC performs best at equal error protection, while Video + Depth is significantly better at unequal error protection. The error protection did not show any impact on the perceived quality at high error rates. An explanation for these results can be found in the fact that UEP allows for better protecting the video view in the Video + Depth approach. The better performance for MVC at EEP can be explained with the additional interview dependencies of left and right view. Taking together the results of the study, Strohmeier et al. conclude that MVC is the strongest coding method contender for mobile 3D media due to its higher error robustness in time-varying wireless channels [122].

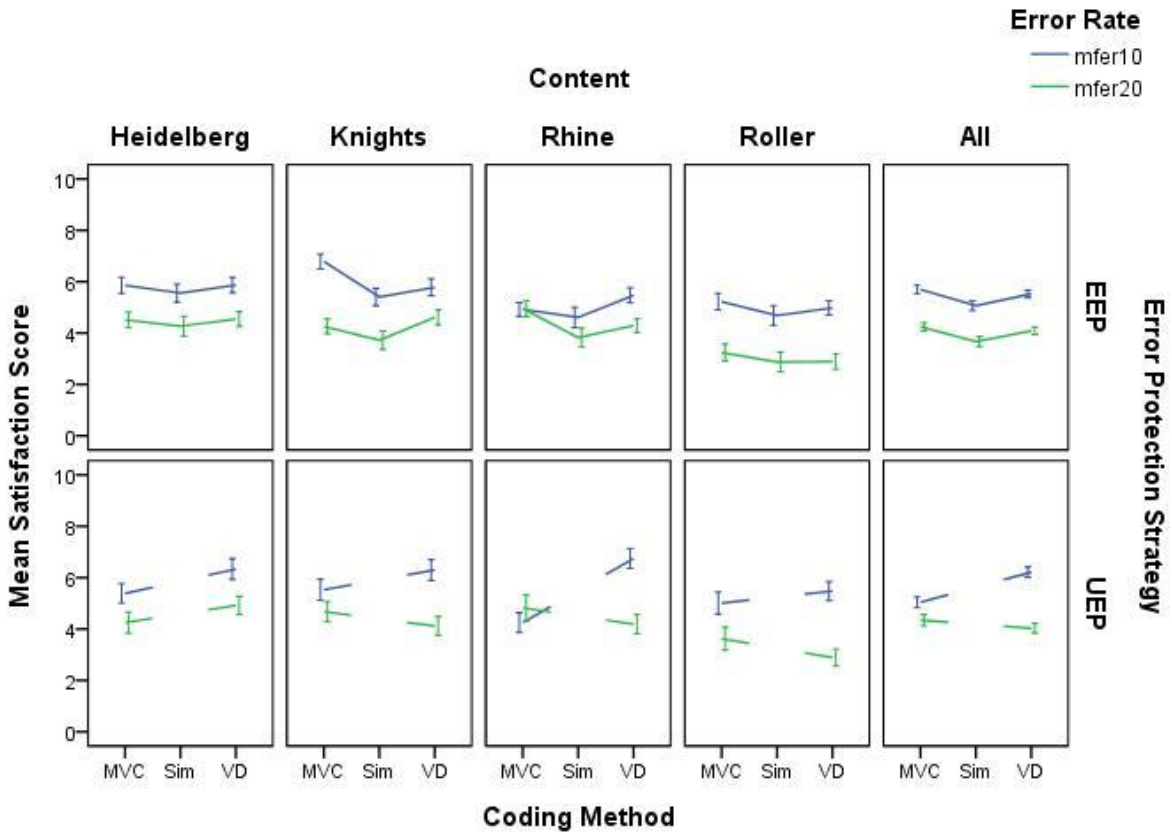


Figure 33. Results of transmission study [122] given as overall results (All) as well as content-per-content. Error bars show 95% CI of mean.

C. User-centered evaluation studies on mobile 3D media

Beyond the quantitative analysis of satisfaction with the overall quality of mobile 3D media systems, the User-Centered Quality of Experience approach [108] targeted a deeper evaluation of the different components that contribute to QoE for mobile 3D media. The application of the Open Profiling of Quality approach [123] resulted in deeper knowledge about the interaction of video quality and depth perception on forming 3D Quality of Experience. In sensory profiling, test participants, in addition to a quantitative profiling, develop their individual quality attributes. These attributes are then used individually to describe the perceived quality. The data is then analyzed using Generalized Procrustes Analysis [113] which results in a low-dimensional and a perceptual model of the experienced quality factors. Two studies by Strohmeier et al. [120], [121] have shown that the video quality and artifact-free video perception is still the key factor for high subjective quality of mobile 3D video. The results, as illustrated in Figure 34, demonstrate that quality mainly depends on one component that has been identified as '*video quality*' as its polarities are described with attributes like mosaic, fuzzy, or grainy on the negative polarity and with sharp, high in contrast, and clear on the positive polarity. Surprisingly enough, a depth-related component has not been identified. Attributes describing depth like *3D reality* or *3dimensional* are included on the positive polarity. These results are in line with previous studies

[110]. The artifact free perception of 3D video is also important in the comparison of mobile 2D vs. 3D video. The added value of 3D is only positively attributed to quality perception if the artifact levels of 2D and 3D stimuli are comparable [120]. According to the findings, subjective quality of current mobile 3D devices is still limited by the quality of the applied displays.

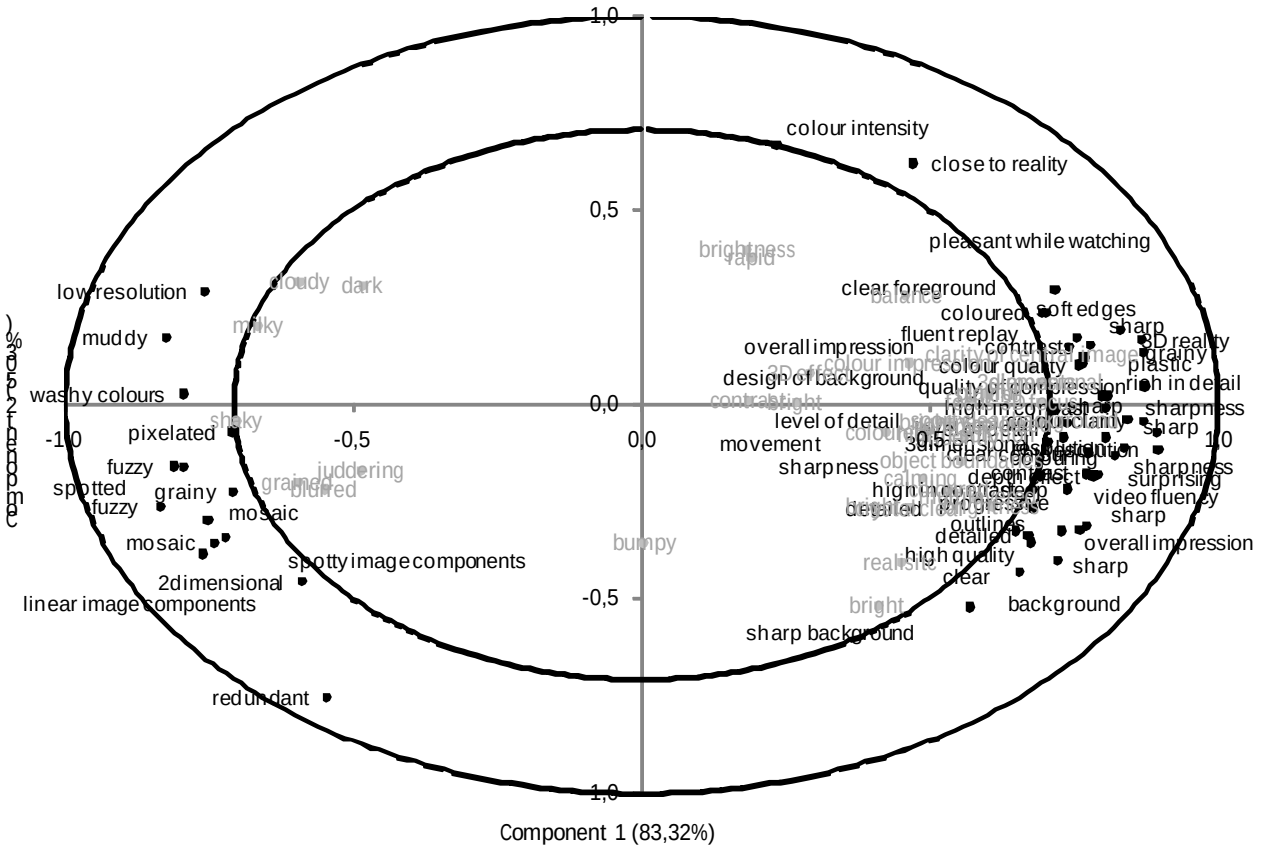


Figure 34. Correlation plot of descriptive study on different coding methods for mobile 3D television and video [121], [123]. Inner circle and outer circle mark 50% and 100% explained variance, respectively.

The collection of individual quality factors from several descriptive studies on mobile 3D media was used by Jumisko-Pyykkö et al. [138] to derive a general set of components of Quality of Experience for mobile 3D video and television systems. The results generalize the individual attributes into four main components of Quality of Experience, namely Visual Quality, Viewing Experience, Content, and Other Modalities Interactions. The component of Visual Quality is further divided into subcomponents of Depth, Spatial, and Motion. Jumisko-Pyykkö et al. [138] provide definitions for each of the components and the underlying factors that contribute to each of these components. Jumisko-Pyykkö et al.'s study transformed the content-dependent findings of the descriptive quality studies [120], [121], [131] into a generalized set of components that can be used in further system developments to guide the design for user-centered quality optimization.

Another main focus in the User-centered Quality of Experience evaluation has been set to the QoE evaluation in the context of use [107] , [131]. It aimed at extending the external validity of the results gained in controlled environments. A recent work on the evaluation of mobile 3D media in the context of use has compared perceived quality in laboratory and different mobile contexts [109], [111], [131]. The work combined quantitative and qualitative evaluation tasks as well as in-depth analysis of contexts and task effort [131]. The results confirm the findings of the user requirements in terms of heterogeneity of the different contexts [116]. Further, the studies have revealed that the results of the quality evaluation differ between controlled environments and the contextual settings. Test participants were less critical in the contextual environments [111]. The studies also showed that quality in the context is depending on the contextual circumstances. Body movements to adjust the viewing distance as well as gaze shifts due to shared attention were significantly higher in the context in comparison to the controlled environment. The strong conclusion is that mobile 3D media systems, beside the 3D experience, need to guarantee ease of viewing as well as a high viewing comfort to provide a high viewing experience in heterogeneous usage contexts [107], [116], [131].

D. 3D graphical user interfaces

It is desirable that the user engage with 3D content actively instead of just being a passive consumer. In addition, the users should also be able to search, browse, and annotate 3D media content, using 3D input modalities. 3D media will benefit from interactivity on mobile devices more than on desktops, because of the limitations of the mobile context, including small physical screen size and limited input modalities. With users demanding ever larger screens and attractive interfaces from mobile devices, the graphical user interface is becoming the most prominent feature of a mobile device.

Several works have studied the creation of 3D interaction techniques that approach the richness of reality, particularly for desktop and large-scale interaction. Shneiderman et al. [58] have examined the features for increasing the usability of 3D user interfaces primarily for desktop and near-to-eye displays, and have proposed general guidelines for 3D UI developers. These guidelines include: better use of occlusion, shadows, and perspective; minimizing the number of steps in navigation in the UI; and improving text readability with better rendering, limited angle to the view position, contrast with the background, and so on. Bowman et al. have analyzed the interaction techniques that are common to applications in 3D User interfaces, and have developed a taxonomy of universal tasks for interacting with 3D environments [59]. These tasks include: selection and manipulation of virtual objects, travel and wayfinding within a 3D environment, issuing commands via 3D menus, and symbolic input such as text, labels, and legends. Defining appropriate 3D interaction techniques is still an active field in itself [59].

3D Widgets

For 3D graphics, however, there is a lack of standardized 3D UI widgets. This is partially due to the lack of commonly accepted list of UI metaphors, and partially due to the lack of an effort to structure a comprehensive and flexible set of existing widgets into a common 3D UI library. Also, when designing 3D user interfaces, new challenges emerge compared to traditional 2D UI design. A major difference between 2D and 3D UIs is how the possibility to position objects in depth (along the z-axis) affects information density. Recent efforts have attempted to standardize a list of 3D widgets [60], [61], [62]. The most popular 3D widgets are based on metaphors that can be listed as tree, card, elevator, gallery, mirror, book, and hinged metaphors [62]. For example, Apple's Coverflow interface that is used in iPhone and Mac OSX Finder applications makes use of the card metaphor.

According to the application and the targeted task, different layout techniques can be selected. Undeniably, depth positioning adds complexity to the design of UIs since more layout options emerge. A stereoscopic 3D UI looks quite different than a 3D UI rendered on a 2D screen. To designers without a lot of prior experience of the characteristics of stereoscopic design, guessing the visual effects of positioning UI elements in depth can be difficult. In Figure 35, 3D graphics is used to display a number of media content for a media browser in a circle seen in different layouts. In Figure 36, another UI example by TAT Inc., called SocialRiver, is shown, where photos, videos, and applications are dropping down in at the far end, and move towards the front. The user can "catch" a photo, video, or application and make it active. This includes showing the video or photo in higher resolution, or activating the application, as shown in the figure. Programmable vertex and pixel shaders are used to render depth-of-field effect and motion blur to direct the focus to the front-most icons, as well as to animate "wobbly" windows using vertex skinning.

Figure 35. Three alternatives for 3D media browser layout [65].

Figure 36. TAT's SocialRiver application.

3D UI performance

In 3D UIs, it is essential to optimize the graphics rendering for power consumption. In stereoscopic rendering, the images for the left and right eyes are very similar, and there is an opportunity to exploit this inherent coherency. With a brute-force implementation, the scene is first rendered to the left eye, and then to the right eye. In general, however, it makes sense to render a single triangle to both views before proceeding with the next triangle [63]. Kalaiah and Capin [64] use this rendering order to reduce the number of vertex shader computations. By splitting the vertex shader into parts that are view-independent (and hence only computed once) and view-dependent, vertex shader computations can be greatly reduced. In the per-pixel processing stage that follows, a simple sorting procedure in a generalized texture space greatly improves the texture cache hit ratio [63], keeping the texture bandwidth very close to that of monoscopic rendering. In addition, Hasselgren and Akenine-Möller [63] introduce approximate rendering in the multi-view pipeline, so that fragment colors in all neighboring views can be approximated from a central view when possible. Otherwise the pipeline reverts to full pixel shader evaluation. When approximate rendering is acceptable, this technique can save a lot of per-pixel shader instructions executions.

To achieve good graphical performance and low power consumption, it is necessary to reduce the internal traffic between the processing elements and the memory. Therefore, mobile graphics solutions make use of data and texture compression to decrease that traffic. This is made even more important with the trend that computation power increases at a faster rate than memory bandwidth. For example, in a recent work, based on the International Technology Roadmap for Semiconductors (ITRS), Owens [12] reports that the processing capability growth is about 71%, while DRAM bandwidth only grows by 25%.

One of the most viable approaches for reducing memory traffic to GPU is compression of textures and buffers [126]. *Textures* can be considered as read-only images which are attached to graphical data. The main requirements of a texture compression/decompression algorithm include fast random access to the texture data, very fast decompression, and inexpensive hardware implementation. The requirement of random access usually implies that a block of pixels is compressed to a fixed size. These requirements have given rise to codecs, such as ETC (Ericsson Texture Compression) and PVRTC (PowerVR Texture Compression), which allow developers to compress textures down to 4 bits per pixel or more without any perceived loss of quality. *Buffers* are different from textures in that they are symmetric: both processes must be performed on hardware in real time. For example, the color buffer can be compressed, and so when a triangle is being rendered to a block of pixels (say, 4×4) in the color buffer, the hardware attempts to compress this block.

Another approach for reducing memory traffic is based on tiling architectures. Tiling architectures are built on the goal to reduce the memory traffic related to frame buffer accesses, which may be one of the costly parts of an application. Instead of storing the full frame buffer in memory, thus transmitting it to the CPU repeatedly during rendering for different objects, only a small tile of the frame buffer is stored on the graphics chip, and the rendering is performed one tile at a time. This approach allows many possibilities for optimization and culling techniques, avoiding processing of data that will not contribute to the final image. Commercially, both Imagination Technologies and ARM provide a mobile 3D accelerator using a tiling architecture.

3D User Input

Utilizing 3D input techniques with autostereoscopic displays provides additional challenges related to the finger occluding the stereo information and virtual buttons being at different depth levels compared to the physical display, as well as problems related to the limited viewing area of the autostereoscopic display. A number of alternatives currently exist on mobile devices for 3D interaction, including the use of touchscreen-based input, inertial trackers, camera-based tracking, GPS tracking – each with its own advantages and disadvantages.

- With *touchscreen-based input,* efficient use of screen space is essential. For single- or multi-touch screen UIs, the main limitation is that interactive elements should be presented in at least 1 x 1 cm square on the touch surface to be picked by an average finger [30]. In return, this limits how many UI elements can be rendered on display. A possible solution is to layer the 3D UI elements, such that the elements in the top layer are large enough to support finger touch input, while rendering can be denser in the underlying layers.

- With *inertial tracker (accelerometer or gyroscope) based input,* there is an advantage that there is no limit on the size of the UI elements. On the other hand, a major problem with inertial trackers is that they suffer from error accumulation due to sensor biases, noise, and drift [47]. In addition, because mobile devices are assumed to be used while on the move, mechanisms are necessary to filter out the jitter created by user's movement (e.g. acceleration due to walking, user in a car) from the user's actual input to the application. Thus, recent research studies have attempted to detect the context from accelerometer input [48].

- *Camera-based input* solutions have also been proposed. Face tracking allows enhancing the depth effect in the applications by supplying motion parallax for enhancing human-computer interaction [48]. In addition, eye-position tracking allows adapting the stereo views to compensate for the zones with high cross-talk and to prevent pseudostereo [140]. Camera input can also be used for tracking the self movement of the device, which can be used for controlling scroll and view direction in an application [50]. Researchers have also proposed a finger tracking system, which allows the user to interact with 3D scenes floating in 3D [55].

V. USE SCENARIOS AND RESEARCH CHALLENGES FOR THE NEXT GENERATION 3D MOBILE DEVICES

3D media-enabled mobile devices are part of a bigger revolution bringing the next generation networked media devices, services and applications where Internet is expected to play the central role. In the incoming years Internet is expected to become larger, faster and more reliable. Its use shall grow beyond simple tasks such as searching for movies or buying food online. The web will evolve from a place for *sharing* content (*Web 2.0*) to a common environment where content is *co-created* (*Web 3.0*) [72]. The media, occupying most of the today's internet – images and video – will evolve to the more realistic, 3D images and 3D video. Consumer electronics will transform from digital (*CE 2.0*) to connected (*CE 3.0*) [73] and will support 3D media as well [74]. Today, most of the 2D media exists in the flat world of web 1.0 and web 2.0 pages. Naturally, the 3D media of tomorrow will "live" on a 3D canvas – **the 3D media internet**. Instead of a proprietary 3D virtual world created by a single organization (such as Second Life [75] or Google Lively [76]), the 3D internet of the future shall be created by its users.

The vision for 3D internet is not brand new. However, earlier attempts – like VRML, X3D and numerous other standards – were not widely accepted by the public. One of the reasons is that creating a VRML model requires too much time and skills compared to shooting a photo and sharing it on Picasa [77] or Flickr [78]. The future 3D internet should allow people to co-create contents and knowledge, and key factor for its success is that the users have tools to create 3D media as simply as it is for making a photo or video today.

We foresee a universal, scalable, user-centric service which will allow 3D media to be co-created and positioned on the 3D continuum of the Future Internet. Such service will combine 3D audio-video data and 3D models both anchored to position in 3D space. In the *3D media cloud* the data will be continuously updated in a recursive manner, as illustrated in Figure 37. Incoming 3D video streams will be used to update the models; the models will be used to register positions in space, which will refine the coordinates of the video streams as well as their 3D quality.

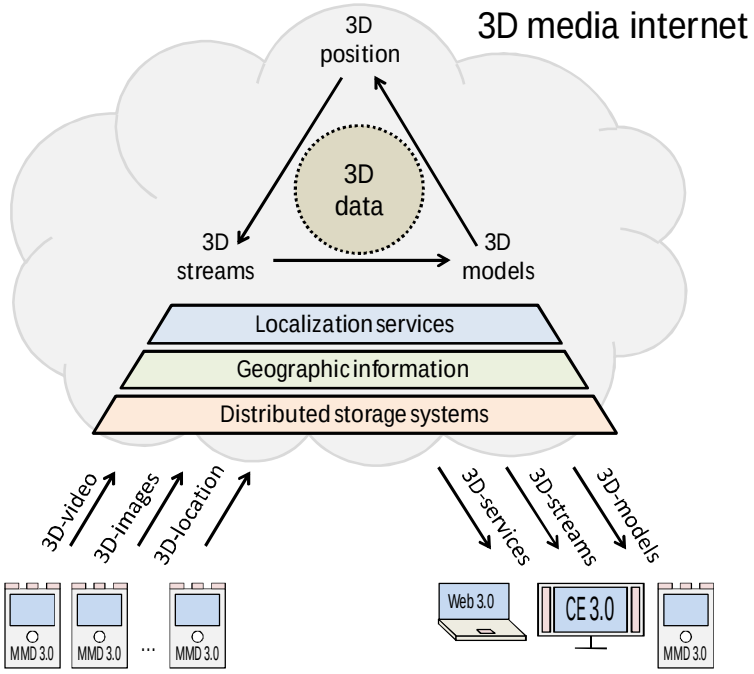


Figure 37. 3D Media Internet concept.

A key element of such service is what we call next-generation 3D-enabled mobile device or *Mobile Multimedia Device 3.0* (*MMD3.0*). It is a portable wireless network terminal, capable of capturing images and video in 3D, recording 3D audio, and being aware of its position and orientation in 3D world. It can capture 3D data, tag it with 3D location and send it to the cloud. It can browse through 3D audio-video streams and 3D models, and visualize them on a 3D display. Many MMD3.0 devices will record data from the 3D world, and sharing this data will create the canvas of the 3D media internet.

The construction of the virtual world will gradually evolve through three stages. At the beginning, user-created content will be roughly positioned on a 2D map. Such services already exist – one example is Google Maps, which relies on

volunteers to create 3D models and position them manually on the map space [79]. Another service, soon to appear is Nokia Image Space [80] where 2D photos are automatically Geo-positioned based on GPS and compass data. Our vision combines both concepts – media will be in 3D and will be automatically Geo-located on a 3D map.

In the second stage, the collected 3D audio-video data will be used to create *3D models* – in the form of point clouds – of the real world. One example for such paradigm is Microsoft Photosynth [81], where 2D photos are used for building rough point-cloud of a scene. The downside of Photosynth is that it requires many 2D images to reconstruct a 3D model, and expects the images to be *manually* tagged as belonging to a certain place. On the contrary, the 3D audio-visual data gathered by an *MMD3.0* type of device will allow reconstruction using much fewer sources. As a result, more precise 3D models will appear at a faster rate in the 3D media internet.

In the third stage, most of the geographic locations in the world will be presented as 3D models. Naturally, the important landmarks will be reconstructed first. As new audio-video streams are available, the 3D models will be continuously updated. The 3D media will appear on the map, and will be available for browsing by location or following hyperlinks. The users might volunteer to improve the quality of the virtual map, since adding new data will be easy "point-and-click" operation. Or, they might contribute by simply sharing their holiday 3D videos. In the 3D media internet, a MMD3.0 compatible device will serve both as distributed sensor network and a terminal. By aiming the device towards a landmark in the real world, it will "know" a) what is in front of the camera and b) the direction of the camera. This will enable services such as 3D location search, 3D position and time-shifting and 3D content browsing and creation.

Challenges

The current architecture of the Internet is progressively reaching a saturation point in meeting increasing user's expectations [72]. Future Internet should be able to grow both in size and throughput to accommodate tomorrow's communication requirements. In order to identify the key research challenges of 3D Media for mobiles we will follow the information flow between users of 3D media and services providing it. Figure 38 illustrates the path of 3D media as it is being captured by and reconstructed on an MMD3.0 device, transmitted over the network, stored in the "cloud", forwarded on request, enhanced and played back on a MMD3.0 network terminal.

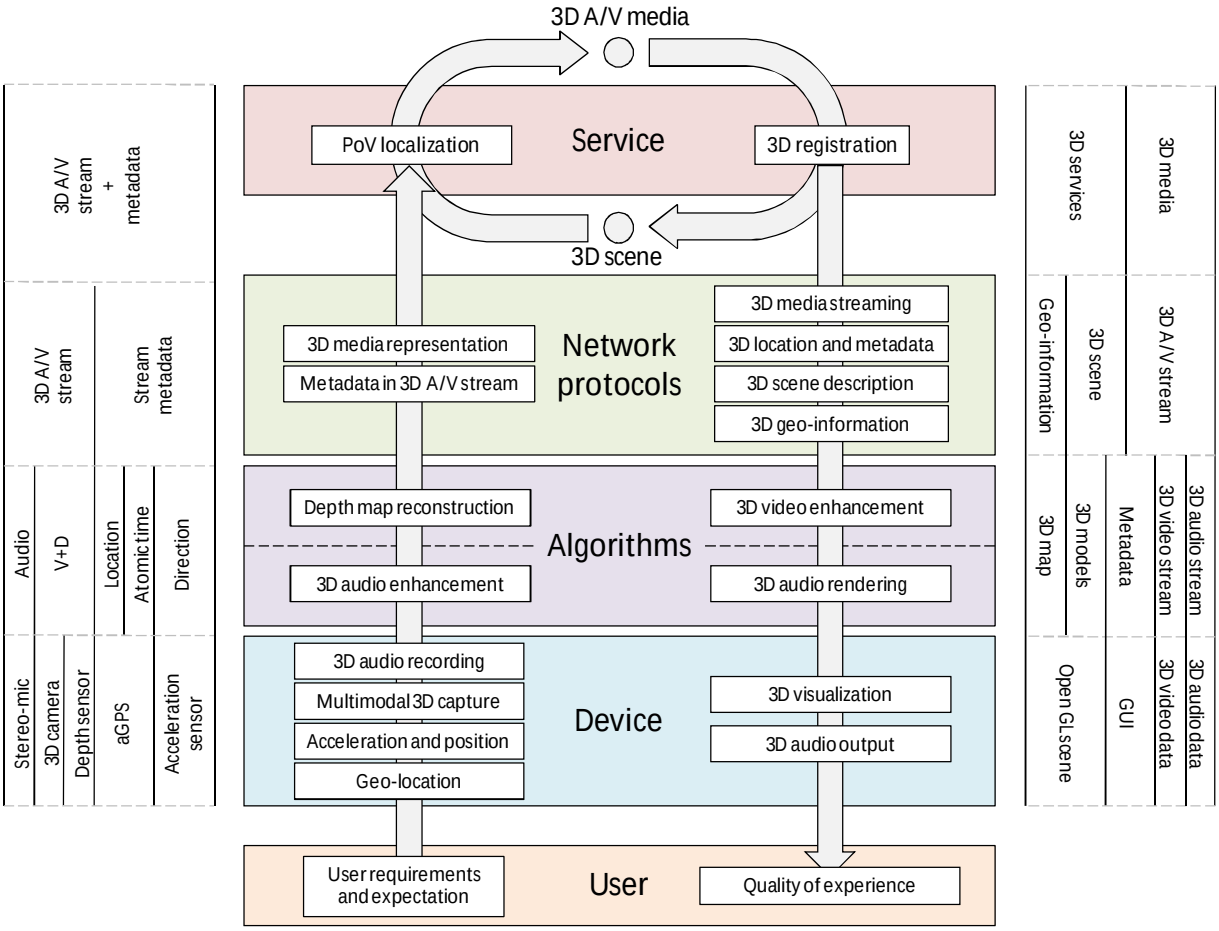


Figure 38. 3D Media path from/to device.

A main research challenge is to make MMD3.0 truly personal. This includes understanding the features and limitations of user-created 3D audio-video content and addressing the quality of experience as perceived by subscribers of 3D media services.

Another research challenge is to seamlessly integrate different sensors for capturing 3D audiovisual information – stereo-microphones, stereo-camera, range sensor, GPS and acceleration sensor along with auto-stereoscopic 3D display and 3D audio output. Sensor fusion and 3D data reconstruction shall be performed in the MMD3.0 terminal. This will require powerful algorithms for converting multisensory data into *3D media content*, i.e. video, augmented with dense depth and location information.

Next challenge is to enable network standards and protocols for representing 3D media as interconnected network objects or in other words, *"3D things"*. Format in which 3D audio-video data, enriched with geographic coordinates, 3D orientation and time-stamps should be defined. On the way back to the MMD3.0 device, it should provide descriptions of 3D scene augmented with 3D objects, as well as 3D media streams. Location, geographic information, and other services provided by the 3D media internet should be requested by and delivered to an MMD3.0 in a scalable

manner. It is also expected that geo-information will become core service of the Future Internet, as search is core service today.

The last challenge in our concept is to deliver 3D media service which contains 3D maps of the real world, libraries of 3D models and 3D audio-video streams located on the maps. This is precisely the stage, when *user-created* content shall become *co-created*. The following functionalities should be supported by such service: 3D models ("point clouds") reconstructed from available 3D streams; position (point-of-view) of MMD3.0 device registered with respect to the models. The "point-of-view localization" and "point cloud reconstruction" tasks shall be re-executed as new 3D media is available, yielding better 3D models and better localization of the 3D media available in the library. According to the "Network tussle" principle [82] several 3D media services in various stages of precision can co-exist in the Future Internet being compatible and standardized. Further research challenges are related with distributed network storage and "cloud computing".

VI. CONCLUSIONS

In this paper, specifics of delivery of 3D media to mobile devices have been addressed. Recent studies have confirmed that the users of such media expect higher realism of and closer emotional relation with the new content. Achieving such realism and emotional effect on a portable device is a challenge both for the optical quality of the display and the methods for creation and delivery of 3D content. To address this challenge in a proper way, the studies of user experience have to be scheduled already at the beginning of the design of the overall system. Furthermore, new methodologies for user studies have to be developed to tackle the complexity of the problems with content, formats, delivery and consumption. Two such methodologies, namely the OPQ and user-centered QoE evaluation in the context of use have been developed and successfully applied with the aim of optimizing new technology components and gathering new knowledge about how users tend to consume 3D content on mobiles. The studies have especially emphasized the importance of visual quality of 3D content for the acceptance of the new technology. The results of these studies have strong implications to the choice of displays, 3D video formats, and coding and transmission methods as well as the receiver-side processing and GUIs.

In the successive stages of development and deployment of 3D services and applications for mobiles, new high quality 3D displays shall be available at first. Portable auto-stereoscopic displays have been the main contender for delivery of 3D visual experience on mobiles. The user studies have elicited the principal characteristics of such displays. They need to be switchable in order to provide the freedom of choosing between 2D and 3D contents and their combination. They need to provide the same quality in 2D and 3D as 3D with decreasing quality is immediately discarded by the user. For such displays the spatial resolution does matter and it should not be compromised for the price of delivering the 3D

effect. Portable 3D displays should guarantee the ease of viewing and ensure high viewing comfort in heterogeneous usage contexts.

After displays, it is the content to be delivered. It is highly determined by the dynamism of mobile users. It should be content for 'fast' consumption: sport events, short documentary, and news. No long watching is expected but 15 to 30 min of use. In addition to television-like content, mobile applications to be used in heterogeneous environments such as interactive navigation and 3D games are highly expected.

3D Video seems the most mature content for mobile delivery. Again, the quality issue is of primary importance, as the user studies revealed that 3D video is accepted as superior to 2D video *only if* artefact-free. This determines the research challenges for the format and coding and transmission methods. Among coding methods, MVC has demonstrated the best rate-distortion performance and robustness in varying channel conditions. These results are consistent with the choice of MVC as the coding format for Blu-Ray. However, this consistency specifies also the next research challenge: how to effectively repurpose high definition 3D content for its mobile visualization as it is expected that 3D video will be mainly created in HD. Simple resizing of stereo video effects in changing the 3D geometry of the scene and diminishing the 3D effect. It seems that there is a need of a genuine master format for 3D video where the depth map of the scene is explicitly presented so to allow a realistic rendering in different perspectives and spatial resolutions.

Precisely because of the demand for high-quality, error protection for robust transmission of 3D video over wireless channels is a must. Optimal combinations of effective coding and effective error protection have been studied especially for the case of DVB-H broadcast and the results have favoured the combination of MVC with application-layer slice-mode error protection and MPE-FEC EEP. Still, UEP approaches bear the potential to achieve higher performance especially if combined with cross-layer optimization.

Along with the quality of 3D content, it is the attractive graphical user interface which must appeal to the mobile users. In contract to content delivery where scalable solutions are likely (i.e. repurposing of HD content, rendering of mobile stereo from multi-view plus multi-depth representations), the graphical user interfaces should be unique and scalable solutions are not possible. Instead, GUIs have to be especially designed for the portable 3D platforms addressing the issue of realism, emotion exploiting the main difference between 2D and 3D – the availability of depth to be used for increasing and enriching the information density.

The first stage of deployment of 3D media to mobiles will be mainly related with media consumption, i.e. delivery of video, GUIs, games. The next stage is to turn the mobile user from a consumer to a creator of 3D content. This would require substantial research efforts, as to make capture in 3D a trivial task. Current state-of-the-art dictates that 3D

capture is highly professional work related with the 3D-specific visual artifacts, which requires a professional planning and shooting combined with post-processing. For mobile 3D capture, these things should be made automatic. In the beginning, mobile 3D camera devices will be with limited quality yet being able to contribute to co-creation of high-quality 3D models and 3D environments, where 3D audio and video, augmented with positioning information will be a basis of novel services and applications.

REFERENCES

[1] A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Müller, C. Erdem, C. Weigel, "Scene Representation Technologies for 3DTV—A Survey", IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1587-1605, Nov. 2007.

[2] A. Smolic, K. Müller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A. Koz, "Coding Algorithms for 3DTV—A Survey", IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1606-1621, Nov. 2007.

[3] G. B. Akar, A. M. Tekalp, C Fehn, M. R. Civanlar, "Transport Methods in 3DTV—A Survey", IEEE Trans. Circuits Syst. Video Technol., vol. 17, pp. 1622-1630, Nov. 2007.

[4] Y.-K. Chen, C. Chakrabarti, S. Bhattacharyya, B. Bougard, 'Signal Processing on Platforms with Multiple Cores: Part 1 – Overview and Methodologies', IEEE Signal Processing Magazine, vol. 26, Nov. 2009, pp. 24-25.

[5] G. Blake, R. Dreslinski, T. Mudge, ' A Survey of Multicore Processors', IEEE Signal Processing Magazine, vol. 26, Nov. 2009, pp. 26-37.

[6] OMAP 4 Platform: OMAP 4430/OMAP4440, available online at http://focus.ti.com/general/docs/wtbu/wtbuproductcontent.tsp?templateId=6123&navigationId=12843&contentId=53243

[7] LH7A400 32-Bit System-on-Chip by NXP Semiconductors, available online at http://www.nxp.com/#/pip/pip=[pip=LH7A400_N_1]|pp=[t=pip,i=LH7A400_N_1]|

[8] Marvel PXA320 Processor Series, available online at http://www.marvell.com/products/cellular/application/pxa320.jsp

[9] NVidia Tegra Product Page, available online at http://www.nvidia.com/page/handheld.html

[10] Next Generation NVIDIA Tegra page, available online at http://www.nvidia.com/object/tegra_250.html

[11] The Snapdragon platform, available online at http://www.qualcomm.com/products_services/chipsets/snapdragon.html

[12]J. D. Owens, Streaming Architectures and Technology Trends. In GPU Gems 2. Addison-Wesley, 457–470, 2005.

[13]S. Pastoor, “3D displays”, in 3D Video Communication, Schreer, Kauff, Sikora, Eds., Wiley, 2005, ch. 13, pp. 235-260.

[14]L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanlar and J. Watson: “An assessment of 3DTV technologies,” NAB Broadcast Engineering Conference Proceedings 2006, pp. 456-467, Las Vegas, USA, April 2006.

[15] C. Van Berkel and J. Clarke, “Characterization and optimization of 3D-LCD module design”, in Proc. Stereoscopic Displays and Virtual Reality Systems IV, SPIE vol. 2653, pp. 179-186, 1997.

[16] W. L. IJzerman, S. T. de Zwart, and T. Dekker, “Design of 2D/3D Switchable Displays,” J. Soc. Inf. Disp. vol. 36, pp. 98-101, May 2005.

[17] Sharp Laboratories of Europe, website, http://www.sle.sharp.co.uk/research/optical_imaging/3d_research.php

[18] S. Uehara, T. Hiroya, H. Kusanagi; K. Shigemura, H. Asada, “1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement”, in Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX, Proc. of SPIE vol. 6803, pp. 68030O-68030O-8 (2008), 2008.

[19] G. J. Woodgate, J. Harrold, “Autostereoscopic display technology for mobile 3DTV applications”, in SPIE-IS&T Electronic Imaging 2007, Stereoscopic Displays and Applications XVIII, Proc. SPIE Vol.6490, pp. 64900K, 2007.

[20] 3D-LCD product brochure, MasterImage, available online at http://masterimage.co.kr/new_eng/data/masterimage.zip?pos=60

[21]H. Hong and M. Lim,”Determination of luminance distribution of autostereoscopic 3D displays through calculation of angular profile”, J. Soc. Inf. Display 18, 327 (2010), DOI:10.1889/JSID18.5.327

[22]M. Salmimaa and T. Jarvenpaa, “3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays”,J. Soc. Inf. Display 16, 1033 (2008), DOI:10.1889/JSID16.10.1033

[23]G.Woodgate and J. Harrold, “Key design issues for autostereoscopic 2-D/3-D displays”, J. Soc. Inf. Display 14, 421 (2006), DOI:10.1889/1.2206104

[24]P. Boher, T. Leroux, V. Collomb Patton, T. Bignon, and D. Glinel, “A common approach to characterizing autostereoscopic and polarization-based 3-D displays”, J. Soc. Inf. Display 18, 293 (2010), DOI:10.1889/JSID18.4.293

[25]F.Kooi, A. Toet, “Visual comfort of binocular and 3D displays”, Displays, Volume 25, Issues 2-3, August 2004, Pages 99-108, ISSN 0141-9382, DOI: 10.1016/j.displa.2004.07.004.

[26]“FinePix REAL 3D series”, product brochure, Fujifilm, 2010, Available: http://fujifilm.co.uk/media/dContent/mediaCentre/Brochures/0_FinePix-Real-3D-catalogue.pdf

[27] J. Konrad and P. Angiel, “Subsampling models and anti-alias filters for 3-D automultiscopic displays”, IEEE Trans. Image Process., vol.15, pp. 128-140, Jan. 2006.

[28] B.A. Wandell, Foundations of Vision, Sinauer Associates, Inc, Sunderland, Massachusetts, USA, 1995.

[29] D. Chandler, “Visual Perception (Introductory Notes for Media Theory Students)”, MSC portal site, University of Wales, Aberystwyth, available at http://www.aber.ac.uk/media/sections/image.html

[30] I. P. Howard and B. J. Rogers, Binocular Vision and Stereopsis, Oxford University Press, New York, Oxford, 1995.

[31] S. Pastoor, “Human factors of 3D imaging: Results of recent research at Heinrich- Hertz- Institut Berlin”, 2nd International Display Workshop, Hamamatsu, pp. 69-72, 1995

[32] D.B. Diner, “A new definition of orthostereopsis for 3-D television”, IEEE International Conference on Systems, Man and Cybernetics, pp.1053-1058, October 1991.

[33] B. Julesz, Foundations of Cyclopean Perception, The University of Chicago Press, Chicago, 1971.

[34] M. Halle, “Autostereoscopic displays and computer graphics”, International Conference on Computer Graphics and Interactive Techniques, 2005.

[35] M. Yuen, “Coding Artefacts and Visual Distortions”, in Digital Video Image Quality and Perceptual Coding, H. Wu. K. Rao, Eds, , CRC Press, 2005.

[36] M. Yuen and H. R. Wu, “A Survey of MC/DPCM/DCT Video Coding Distortions,” Signal Processing, vol. 70, pp. 247–278, Nov. 1998.

[37] M. McCauley and T. Sharkey, “Cybersickness: Perception of Self-Motion in Virtual Environments”, Presence: Teleoperators and Virtual Environments, vol. 1, pp. 311-318, 1992.

[38] M. Wexler and J. Boxtel, “Depth perception by the active observer“, Trends in Cognitive Sciences, vol. 9, pp. 431-438, Sept, 2005.

[39] A. Boev, D. Hollosi, and A. Gotchev, “Classification of stereoscopic artefacts”, MOBILE3DTV Tech. report D5.1, 2008.

[40] W. IJsselsteijn, P. Seuntiens and L. Meesters, “Human factors of 3D displays”, in 3D Video Communication, Schreer, Kauff, Sikora, Edts., Wiley, 2005.

[41] C. Lin, C. Ke, C. Shieh, and N. K. Chilamkurti, “The packet loss effect on MPEG video transmission in wireless networks”, in Proc. 20th int. Conference on Advanced information Networking and Applications - (Aina'06), vol. 1. pp. 565-572, 2006.

[42] P. Surman, K. Hopf, I. Sexton, W.K. Lee, R. Bates, “Solving the 3D problem - The History and Development of Viable Domestic 3-Dimensional Video Displays”, In Three-Dimensional Television: Capture, Transmission, and Display, H. M. Ozaktas, L. Onural, Eds., (ch. 13), Springer Verlag, 2007.

[43] M. Hassenzahl and N. Tractinsky, “User Experience – a Research Agenda.” Behaviour and Information Technology, Vol. 25, No. 2, pp. 91-97, March-April 2006.

[44] V. Roto, “Web Browsing on Mobile Phones – Characteristics of User Experience.” Doctoral dissertation, TKK Dissertations 49, Helsinki University of Technology, Finland, 2006.

[45] D. Saffer, Designing Gestural Interfaces: Touchscreens and Interactive Devices, O'Reilly Media, Inc., 2008.

[46] S. Mizobuchi et al., “The Effect of Stereoscopic Viewing in a Word-Search Task with a Layered Background,” Journal of the Society for Information Display, vol. 16, pp.1105-1113.

[47] E. Foxlin, “Motion tracking requirements and technologies”. In K. M. Stanney (Ed.), Handbook of virtual environ-ments: Design, implementation, and applications (p. 167).NJ: Erlbaum

[48]L. Bao, S. Intille, “Activity recognition from user-annotated acceleration data”, in Proc. PERVASIVE 2004, Springer LNCS 3001, pp. 1-17, 2004.

[49] M. Barnard, J. Hannuksela, P. Sangi, J. Heikkilä, “A vision based motion interface for mobile phones”, in Proc. 5th International Conference on Computer Vision Systems (ICVS), Bielefeld, Germany, 2007.

[50] T. Capin, A. Haro, V. Setlur, S. Wilkinson, “Camera-Based Virtual Environment Interaction on Mobile Devices”, Lecture Notes in Computer Science. vol. 4263, pp. 765-773, Springer, 9783540472421, Berlin, 2006.

[51] J. Hannuksela, P. Sangi, J. Heikkilä, “A Vision-Based Approach for Controlling User Interfaces of Mobile Devices”, in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Vision for Human-Computer Interaction (V4HCI). vol. 6 pp. 71, San Diego, CA, (2005)

[52] B. Hjelmas, “Face Detection: A Survey”, Computer Vision and Image Understanding. Vol. 3, pp. 236-274 (2001).

[53] G. R. Bradski, “Computer vision face tracking for use in a perceptual user interface”. Intel Technology Journal. 95-103 (1998).

[54] A. Bulbul, Z. Cipiloglu, T. Capin, “A Color-Based Face Tracking Algorithm for Enhancing Interaction with Mobile Devices”, The Visual Computer Journal, Springer, 2010.

[55] J. Hannuksela, S. Huttunen, P. Sangi, J. Heikkilä, “Motion-based Finger Tracking for User Interaction with Mobile Phones”, in Proc. of 4th European Conference on Visual Media Production (CVMP). London, UK, (2007).

[56]T. Höllerer, “User interface management techniques for collaborative mobile augmented reality”, Computers & Graphics vol. 25, pp. 799-810, 2001.

[57] S. Bhandari, Y. K. Lim, “Exploring Gestural Mode of Interaction with Mobile Phones”, Proc. CHI 2008 – Works in Progress, pp. 2979-2984, 2008.

[58] B. Shneiderman, C. Plaisant, Designing the User Interface, Fourth Edition, Addison-Wesley, 2004.

[59] D. Bowman, E. Kruijff, J. J. LaViola, I. Poupyrev, 3D User Interfaces – Theory and Practice, Addison-Wesley, 2004.

[60] R. Dachselt, A. Hubner, “Virtual environments: Three-dimensional menus: A survey and taxonomy,” Computers & Graphics, vol. 31: pp. 53–65, 2007.

[61] 3D Components Web Site www.3dcomponents.org.

[62] EC FP7 3DPHONE Project "Interim Progress Report for 3D UI / Direct Manipulation Solutions," Tech. Rep. D6.1.2 (2009).

[63] J. Hasselgren, T. Akenine-Möller, “An Efficient Multi-View Rasterization Architecture”, in Proc. Eurographics Symposium on Rendering, 61–72, 2006.

[64] A. Kalaiah, T. Capin, “Unified Rendering Pipeline for Autostereoscopic Displays”, in Proc. 3DTV Conference 2007, May 2007, Kos, Greece.

[65] TAT Inc. "Application built on 3D rendering & 3D display, joystick interaction," 3DPHONE Project Tech. Rep. D6.3.3 (2009).

[66] J. Poikonen, J. Paavola, “Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer”, Proc. ICC 2006, Istanbul, Turkey, 2006.

[67] COST207, Digital land mobile radio communications (final report), Commission of the European Communities, Directorate General Telecommunications, Information Industries and Innovation, 1989, pp. 135 147

[68] G. Akar, M. Oguz Bici, A. Aksay, A. Tikanmäki, and A. Gotchev, “Mobile stereo video broadcast”, Mobile3DTV Project report D3.2, available online at http://sp.cs.tut.fi/mobile3dtv/results/tech/D3.2_Mobile3DTV_v1.0.pdf

[69] T. Marc, M. Lambooij, W. IJsselsteijn and I. Heynderickx, “Visual discomfort in stereoscopic displays: a review”, in Procedings of the SPIE- IS&T Electronic Imaging Vol. 6490, Stereoscopic Displays and Virtual Reality Systems XIV, 5 March 2007.

[70] A. J. Woods, T. Docherty and R. Koch, “Image distortions in stereoscopic video systems”, in Proc. SPIE Vol. 1915, Stereoscopic Displays and Applications, San Jose, California, 23 September 1993.

[71] M. J. Meesters, W. A. IJsselsteijn and P. J. Seuntiens, “A survey of perceptual evaluations and requirements of three-dimensional TV”, in IEEE Trans. Circuits and Syst. Video Technol., vol. 14, pp. 381 – 391, March 2004.

[72] “Future Internet: The Cross-ETP Vision Document”, European Future Internet Portal, available online at http://www.future-internet.eu/fileadmin/documents/reports/Cross-ETPs_FI_Vision_Document_v1_0.pdf

[73] “The Digital Home Experience – Consumer Electronics 3.0”, Intel, Available online at http:// http://www.intelconsumerelectronics.com/Consumer-Electronics-3.0/

[74] L. Onural, “Television in 3-D: What Are the Prospects?”, Proceedings of the IEEE, vol. 95, pp. 1143-1145, June 2007.

[75] Second Life, available at http://secondlife.com

[76] Lively by Google, available at http://www.lively.com

[77] Picasa by Google, available at http://picasa.google.com

[78] Flickr by Yahoo, available at http://flickr.com

[79] Google SketchUp, available at http://sketchup.google.com

[80] Nokia Image space, available at http://research.nokia.com/imagespace

[81] Photosynth by Microsoft Live Labs, available at http://livelabs.com/photosynth/

[82] D. Clark, “Tussle in cyberspace: defining tomorrow's internet”, IEEE/ACM Trans. Netw., vol. 13, pp. 462-475 (2005).

[83] T. Wiegand, G.J. Sulivan, G. Bjøntegaard,, & A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, IEEE Trans. Circuits Syst. Video Technol., vol. 13, July 2003, pp. 560-576.

[84] A. Aksay, G. Bozdagi Akar, "Evaluation of Stereo Video Coding Schemes for Mobile Devices", *3DTV-CON 2009*, Potsdam, Germany, May 2009.

[85] P. Merkle, H. Brust, K. Dix, Y. Wang, A. Smolic, ‘Adaptation and optimization of coding algorithms for mobile 3DTV’, Mobile3DTV tech. report D2.2, November 2008.

[86] Aljoscha Smolic , Karsten Mueller , Philipp Merkle , Peter Kauff , Thomas Wiegand, An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution, Proceedings of the 27th Picture Coding Symposium, p.389-392, May 06-08, 2009, Chicago, IL, USA

[87] A. Vetro, W. Matusik, H. Pfister, J. Xin, “Coding approaches for end-to-end 3D TV systems,” Picture Coding Symposium, 2004.

[88]Video Group: Text of ISO/IEC 14496-10:2009/FDAM1: Constrained baseline profile, stereo high profile and frame packing arrangement SEI message. ISO/IEC JTC1/SC29/WG11 Doc. N10707, London, UK (2009).

[89] H. Lee, K. Yun, N. Hur, J. Kim, Bu C. Min, and J. K. Kim, “A structure for 2D/3D mixed service based on terrestrial DMB system”, in Proc. 3DTV Conference 2007, Kos Island, Greece, 2007.

[90] B. Furht and S. Ahson. Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T, and MediaFLO. Auerbach Publications, 2008.

[91] S. Cho, N. Hur, J. Kim, K. Yun, and S. Lee, “Carriage of 3D audio-visual services by T-DMB,” Proc ICME, 2006.

[92] http://www.3dphone.org/. 3dphone project.

[93] www.mobile3dtv.eu. Mobile3dtv: Mobile 3dtv content delivery over DVB-H system.

[94] cordis.europa.eu

[95]S. Wenger, M. M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, “RFC 3984: RTP payload format for H.264 video.” [Online]. Available: http://tools.ietf.org/html/rfc3984.

[96] M. O. Bici, A. Aksay, A. Tikanmaki, A. Gotchev, and G. Bozdagi Akar, “Stereo Video Broadcasting Simulation for DVB-H”, NEM-Summit'08, 2008.

[97] ETSI, Digital Video Broadcasting (DVB): DVB-Himplementation guidelines, 2009. TR 102 377 V1.3.1.

[98]www.dvb.org/groups_modules/commercial_module/cm3dtv/

[99] D. Bugdayci, M. Oguz Bici, A. Aksay, M. Demirtas, G. B. Akar, A. Tikanmäki, A. Gotchev, “Stereo DVB-H broadcasting system with error resilient tools”, Mobile3DTV Technical report, D3.4, February 2010, available online at: http://sp.cs.tut.fi/mobile3dtv/results/tech/D3.4_Mobile3DTV_v1.0.pdf

[100] S. Worrall, A. H. Sadka, P. Sweeney, and A. M. Kondoz, “Prioritisation of data partitioned MPEG-4 video over mobile networks”. ETT-European Transactions on Telecommunications. Vol.12-3. May/June 2001.

[101] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," IEEE Trans. Circuits and Syst. Video Technol., vol.17, no.11, pp.1461-1473, Nov. 2007.

[102] M. Kornfeld, "Optimizing the DVB-H Time Interleaving Scheme on the Link Layer for High Quality Mobile Broadcasting Reception," IEEE International Symposium on Consumer Electronics (ISCE), Dallas, USA, Jun. 2007.

[103] R. Calonati, “Techniche a diversita compatiili con il livello fisico dello standard DVB-T/H”, Laurea in Ingengeria thesis, Universita degli studi di Firenze, 2007.

[104] S. Jumisko-Pyykkö, J. Häkkinen, G. Nyman, “Experienced Quality Factors – Qualitative Evaluation Approach to Audiovisual Quality”. Proceedings of IST/SPIE conference Electronic Imaging, Multimedia on Mobile Devices, 2007.

[105] S. Jumisko-Pyykkö, and M. Hannuksela, “Does Context Matter in Quality Evaluation of Mobile Television?". In the Proceedings of 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2008)

[106] S. Jumisko-Pyykkö, and D. Strohmeier, “Report on research methodologies for the experiments”, MOBILE3DTV Technical report D4.2, available online at http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.2_Mobile3dtv_v2.0.pdf .

[107] S. Jumisko-Pyykkö, and T. Vainio, “Framing the Context of Use for Mobile HCI”, review paper about mobile contexts of use between 2000-2007, International Journal of Mobile-Human-Computer-Interaction (IJMHCI), in press.

[108] A. Gotchev, A. Smolic, S. Jumisko-Pyykkö, D. Strohmeier, G.B. Akar, P. Merkle, N. Daskalov, "Mobile 3D television: Development of core technological elements and user-centered evaluation methods toward an optimized system". Proceedings of IST/SPIE Conference on Electronic Imaging, Vol. 7256, 3D Video Delivery for Mobile Devices (2009);

[109] S. Jumisko-Pyykkö, T. Utriainen, ”User-centered quality of experience of mobile 3DTV: How to evaluate quality in the context of use.” Proc. SPIE, Vol. 7542, 75420W (2010); doi:10.1117/12.849572

[110] W. Tam, L. Stelmach, and P. Corriveau, “Psychovisual aspects of viewing stereoscopic video sequences," Proceedings of the SPIE 3295, pp. 226-235, 1998.

[111] S. Jumisko-Pyykkö, T. Utriainen, “User.centered Quality of Experience: Is mobile 3D video good enough in the actual context of use?” Proc. VPQM 2010, Scottsdale, AZ, USA, 2010.

[112] S. Jumisko-Pyykkö, V. Kumar Malamal Vadakital, M. Hannuksela, “Acceptance Threshold: Bidimensional Research Method for User-Oriented Quality Evaluation Studies.” Int. Journal of Digital Multimedia Broadcasting, 2008.

[113] J. Gower, “Generalized procrustes analysis,” Psychometrika, vol. 40, 33-51, 1975.

[114] Recommendation ITU-R BT.500-11. 2002. Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-11. ITU Telecom. Standardization Sector of ITU.

[115] Recommendation ITU-T P.910. 1999. Subjective video quality assessment methods for multimedia applications, Recommendation ITU-T P.910. ITU Telecom. Standardization Sector of ITU.

[116] S. Jumisko-Pyykkö, M. Weitzel, D. Strohmeier, “Designing for User Experience – What to expect from mobile 3D television and video”, in Proc. uxTV 2008, Mountain View, CA, USA, October 2008

[117] D. Strohmeier, S. Jumisko-Pyykkö, M. Weitzel, S. Schneider, “Report on User Needs and Expectations for Mobile Stereo-video,” Mobile3DTV report D4.1, available online at http://sp.cs.tut.fi/mobile3dtv/results/tech/D4.1_Mobile3DTV_v1.0.pdf

[118] ISO 13407. 1999. Human-centred design processes for interactive systems. International Standard, the International Organization for Standardization.

[119] J. Freeman, S. E. Avons, “Focus Group Exploration of Presence through Advanced Broadcast Services”, in Proc of SPIE, Human Vision and Electronic Imaging, pp.3959-76, 2000.

[120] D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, “New, lively, and exciting or just artificial, straining, and distracting? A sensory profiling approach to understand mobile 3D audiovisual quality,” in Proc. 4th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM, Scottsdale, USA, Jan. 2010.

[121] D. Strohmeier, and G. Tech, “Sharp, bright, three-dimensional: open profiling of quality for mobile 3DTV coding methods,” in Proc. “Multimedia on Mobile Devices” as part of the SPIE Electronic Imaging Conf. 2010, Multimedia on Mobile Devices at Electronic Imaging 2010, San Jose, California, USA, Jan. 2010

[122] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, G. Tech, D. Bugdayci, and M. O. Bici, “Results of quality attributes of coding, transmission and their combinations,” Technical Report Mobile 3DTV, Jan. 2010

[123] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, “Open Profiling of Quality: A mixed method approach to understand multimodal quality perception”, Advances in Multimedia, vol. 2010, Article ID 658980, 28 pages, 2010. doi:10.1155/2010/658980.

[124] D. Strohmeier, G. Tech, “On comparing different codec profiles of coding methods for mobile 3D television and video”, in Proceedings of the 2nd conference on 3D Systems and Applications 3DSA, Tokyo, Japan, May 2010.

[125] H. Stone, and J. L. Sidel, Sensory evaluation practices , 3rd ed. ed. Academic Press, San Diego, 2004.

[126] T. Capin, K. Pulli, and T. Akenine-Möller, ”The State of the Art in Mobile Graphics Research,” IEEE Comput. Graph. Appl. Vol. 28, pp. 74-84, Jul. 2008.

[127] H. Brust, A. Smolic, K. Müller, G. Tech, T. Wiegand, “Mixed resolution coding of stereoscopic video for mobile devices,” 3DTV Conference 2009, Potsdam, May 2009.

[128] H. Brust, G. Tech, K. Müller, T. Wiegand, “Mixed resolution coding with interview prediction for mobile3DTV”, in Proc. 3DTV Conference 2010, Tampere, Finland, June 2010 .

[129] ITU-T Recommendation P.10 Amendment 1, “Vocabulary for performance and quality of service. New appendix I Definition of Quality of Experience (QoE)”, International Telecommunication Union, Geneva, Switzerland, 2008.

[130] Escofier, B., Pagès, J. “Multiple factor analysis (AFMULT package)”, Computational Statistics & Data Analysis, Volume 18, Issue 1, August 1994, Pages 121-140, ISSN 0167-9473, DOI: 10.1016/0167-9473(94)90135-X

[131] Jumisko-Pyykkö, S., Utriainen, T. “A Hybrid Method for Quality Evaluation in the Context of Use for Mobile (3D) Television”, Multimedia Tools and Applications, 2010, in press

[132] Wu, W., Arefin, A., Rivas, R., Nahrstedt, K., Sheppard, R., and Yang, Z. 2009. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework. In Proceedings of the Seventeen ACM international Conference on Multimedia (Beijing, China, October 19 - 24, 2009). MM '09. ACM, New York, NY, 481-490. DOI= http://doi.acm.org/10.1145/1631272.1631338

[133] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), ITU-T and ISO/IEC JTC 1, Advanced Video Coding for Generic Audiovisual Services (November 2007).

[134] ISO/IEC JTC1/SC29/WG11, Text of ISO/IEC 14496-10:200X/FDAM 1 Multiview Video Coding. Doc. N9978, Hannover, Germany (July 2008).

[135] ISO/IEC JTC1/SC29/WG11, ISO/IEC CD 23002-3: Representation of auxiliary video and supplemental information. Doc. N8259, Klagenfurt, Austria (July 2007).

[136] Uehara, S., Hiroya, T., Kusanagi, H., Shigemura, K., and Asada, H. “1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement,” in Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX, Vol. 6803, San Jose, USA, Jan. 2008.

[137] Bugdayci, D., Bici, M.O., Aksay, A., Demirtas, M., Akar, G.B., Tikanmäki, A., Gotchev, A., “Stereo DVB-H broadcasting system with error resilient tools”, Technical report D3.4, December 2009

[138] Jumisko-Pyykkö, S., Strohmeier, D., Utriainen, T., Kunze, K. “Descriptive Quality of Experience for Mobile 3D Video”, to be presented at the 6th Nordic Conference on Human-Computer Interaction (nordiCHI2010), Reykjavik, Iceland

[139] Abbott, B. J. 1986. An Integrated Approach to Software Development. New York: John Wiley.

[140] A. Boev, M. Georgiev, A. Gotchev, N. Daskalov and K. Egiazarian “Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display”, in Proc. of 17th European Signal Conference EUSIPCO 2009, Glasgow, Scotland, August 2009.

[141] Mobile Broadcast/Multicast Service (MBMS), TeliaSonera White paper, available online at http://www.medialab.sonera.fi/workspace/MBMSWhitePaper.pdf

[142] A. Bourge, J. Gobert, and F. Bruls, 'MPEG-C Part 3: Enabling the introduction of video plus depth contents', Proceedings of the Workshop on Content generation and coding for 3D-television, Eindhoven University of Technology, 2006, http://vca.ele.tue.nl/events/3Dworkshop2006/proceedings.html

[143] L. Pasquier, J. Gobert, "Multi-view renderer for auto-stereoscopic mobile devices," IEEE International Conference on Consumer Electronics, 2009 Digest of Technical Papers, 2009, p. 1-2.

[144] ISO/IEC JTC1/SC29/WG11, Vision on 3D Video, Doc. N10357, Lausanne, Switzerland - February 2009.

[145] ST Ericsson: U8500 - The best smartphone platform, available online at: http://www.stericsson.com/platforms/U8500.jsp

[146] J. Konrad, B. Lacotte, E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," IEEE Transactions on Image Processing, vol.9, no.5, pp.897-908, May 2000 doi: 10.1109/83.841535

[147] N. Atzpadin, P. Kauff, and O. Schreer. "Stereo Analysis by Hybrid Recursive Matching for Real-Time Immersive Video Conferencing", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Immersive Telecommunications, 14(3):321-334, March 2004.

[148] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, R. Tanger, Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability, Signal Processing: Image Communication, Volume 22, Issue 2, February 2007, Pages 217-234.

[149] A. Boev, M. Poikela, A. Gotchev, A. Aksay, 'Modeling of the stereoscopic HVS', Mobile3DTV Tech. report D5.3, 2010.

This page left blank

[P03] A. Boev, R. Bregovic, A. Gotchev, "Methodology for design of antialiasing filters for autostereoscopic displays", Special issue on *Advanced Techniques on Multirate Signal Processing for Digital Information Processing, Journal of IET Signal Processing,* Volume 5, Issue 3, June 2010, pp. 333-343

This page left blank

Methodology for design of antialiasing filters for autostereoscopic displays

Atanas Boev, Robert Bregovic, Atanas Gotchev
Department of Signal Processing, Tampere University of Technology, Tampere, Finland
firstname.lastname@tut.fi

ABSTRACT

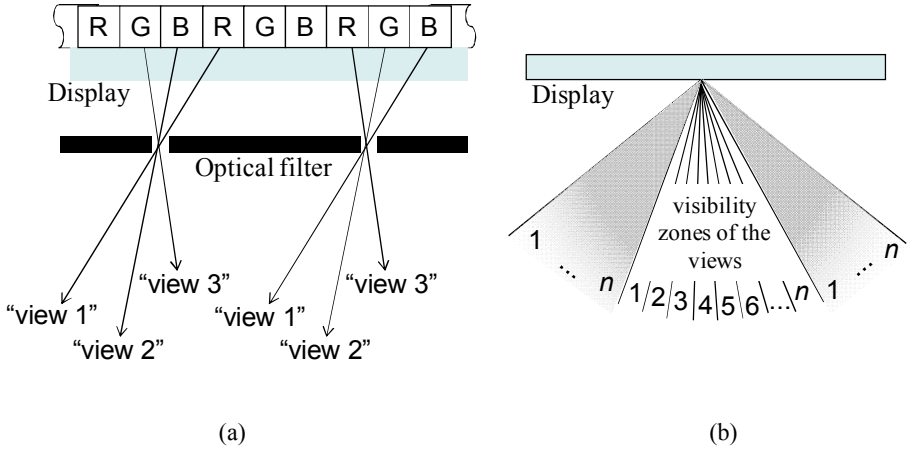
Multi-view autostereoscopic displays can be considered as a kind of a multirate system due to the construction compromise between the number of different views and spatial resolution adopted for such displays. Images to be visualized on these displays are prone to aliasing errors. Careful antialiasing requires knowledge about the display frequency response, which is determined mainly by the view sub-sampling topology but it is also influenced by some other, generally nonlinear, aliasing-causing display effects. In this work, a methodology for designing antialiasing filters for autostereoscopic displays is proposed. It includes the following three steps: 1) measuring aliasing effects by a set of test images – displayed on the screen, then photographed and then analyzed in Fourier domain; 2) estimating the display passband based on the set of measurements; and 3) designing filters confined to the so-estimated band. Using this methodology, one non-separable and three separable antialiasing filters have been designed. The non-separable filter cancels the aliasing terms completely, while the separable approximations allow for some small amount of aliasing for the sake of perceptually-favored sharpness preservation. The advantage of the methodology with respect to previously suggested antialiasing filter design approaches is demonstrated by objective comparisons of filter performance and computational efficiency and by visual inspection on a set of test images.

1. INTRODUCTION

In the recent years, a new group of 3D displays, referred to as *autostereoscopic displays*, has emerged. Such displays create illusion of depth by delivering separate images to each observer's eye without a need for special glasses. Instead, they operate by redirecting the light coming from pixels of a conventional TFT-LCD to different directions thus forming two or multiple views. The effect of redirecting the light is achieved by an optical filter, mounted on top of the LCD surface [1], [2], [3]. There are two common types of optical filters – lenticular sheet [1] which works by refracting the light, and parallax barrier [3] which works by blocking the light in certain directions.

Conventional TFT displays recreate full colour range by emitting light though red, green and blue coloured components (*sub-pixels*). In autostereoscopic mode, sub-pixels appear displaced in respect to the optical filter, and their light is redirected towards different positions. The effect is illustrated in Figure 1(a). The image, formed by the set of sub-pixels, visible from given direction is said to form a *view* [1], [2]. For each view, there is an *optimal observation spot*, where the view is perceived with maximum brightness. The range of angles from which a view can be seen, even though with diminished brightness, is known as the *visibility zone* of that view. Usually, the visibility zones of all views

appear in horizontal direction in front of the display, as depicted in Figure 1(b). In order to visualize a scene in 3D, each view should represent different observation of that scene. The process of mapping an image to the sub-pixels corresponding to one view is called *view interleaving* [1], [4]. The map of correspondences between addressable sub-pixels of the display and the view they belong to is called *interleaving map*. Usually the interleaving map has repetitive structure, which can be represented by an *interleaving pattern* copied multiple times over the display surface. In order to balance horizontal versus vertical resolution of each view, the optical filter is mounted at a slant to the TFT matrix. As a result, the sub-pixels visible from certain observation appear on a non-rectangular grid, which follows the slant of the optical filter. An example of such grid is given by the sub-pixels marked with "F" in Figure 2(a).



(a) (b)

Figure 1, Operation principles of multi-view displays: a) Light redirection by optical filter, b) visibility zones

By moving laterally in front of a multi-view display, one can notice the discrete structure of the views seeing particular types of artefacts. One is *image flipping*, caused by the noticeable transition between the viewing zones, and the other is *picket fence effect*, caused by the gaps between sub-pixels being predominantly visible for some observation angles. The common practice to mitigate these effects is to broaden the visibility angle of each view, thus interspersing the visibility zones [1]. Thus, for any observation angle, a number of views are simultaneously visible, as exemplified in Figure 2(b). To the view originally intended to be visible with full brightness ("F" sub-pixels), views in the neighbouring zones seen with partial brightness are added, as denoted by "P". This effect can be regarded as *inter-channel crosstalk* [1], or *interperspective aliasing* [4].

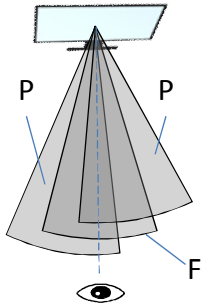
When a 3D object is visualized on a multi-view display with n views, n different observations are interleaved into one compound 3D image. A flat 2D object, which is meant to appear floating in front or behind the screen surface, is represented by n identical observations. In this case, the optical filter can be regarded as a mask, which partially covers the underlying 2D image, or equivalently as a sub-sampling function applied to it. The slanted optical barrier introduces artefacts to the underlying image, which are modelled as aliasing. These artefacts are especially pronounced in flat, two-dimensional parts of the image. Graphical elements of the user interface, movie subtitles and 2D photographs are some

examples of objects, prone to aliasing. Furthermore, aliasing artefacts are most pronounced for a static observer, since the group of visible pixels remains unchanged.

The effect of viewing simultaneously sub-pixels intended to be visible (those marked “F” in Figure 2(a)) with those, which belong to adjacent views (marked “P” in the same figure) has to been taken into account when modelling the sampling pattern. It is not a perfect binary mask but a more comprehensive masking function, where the visibility of each sub-pixel is affected by its relative position in respect to the optical filter.

	Column 1			Column 2			Column 3			Column 4			Column 5			
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R
ROW 1			P	F	P						P	F	P			
ROW 2		P	F	P						P	F	P				
ROW 3		P	F	P						P	F	P				
ROW 4	P	F	P						P	F	P					
ROW 5	F	P						P	F	P						P
ROW 6	F	P						P	F	P						P
ROW 7	P						P	F	P						P	F
ROW 8						P	F	P						P	F	P
ROW 9						P	F	P						P	F	P
ROW 10					P	F	P						P	F	P	
ROW 11				P	F	P						P	F	P		
ROW 12				P	F	P						P	F	P		

(a)



(b)

Figure 2, Aliasing on multi-view displays: a) visibility of sub-pixels, b) interspersing of visibility zones

Research on antialiasing filters for multi-view displays is rather scarce. Jain and Konrad [5] introduced a method for designing 2D non-separable antialiasing filters for an arbitrary sub-sampling pattern. They devised an optimization procedure targeting 2D filter with passband that spans all frequencies at which the contribution of all alias terms is smaller than the original signal itself. In [6], Moller and Travis used simplified optical filter model to analyse display bandwidth, and derive a spatially-varying 2D filter which requires knowledge of scene per-pixel depth. Zwicker *et al.* [4] proposed a low-pass filter to be applied on the sampling grid of the multi-view display expressed in ray-space. Their approach aims at preventing both intra- and inter-perspective aliasing. However, their model assumes constant (i.e. vertical) masking pattern for each image row, which does not take into account the directionally dependant aliasing caused by slanted optical filter. In [7] the authors have proposed a methodology for designing optimal antialiasing filters, based on subjective preferences of the observer.

This paper presents an approach for designing optimal antialiasing filters based on objective criteria. Ideally, a precise model of the optical filter would allow optimal design. However, such model depends on optical parameters of the display which are rarely provided to the owner. An alternative approach, described in this paper, is to directly measure the relevant optical properties of a multi-view display, and use the measurement results for designing the optimal antialiasing filter.

The paper is organized as follows: Section 2 gives insights about the way optical parameters of a multi-view display can be estimated, and presents measurement results for a particular multi-view display, regarded as a case study. In Section

3, two different methodologies for designing antialiasing filters optimized for a specific 3D display are presented. First, the measurement results are used for creating a model and designing non-separable 2D antialiasing filter for the considered display. Then, an attempt is done to reproduce comparable results using less-computationally demanding separable filters. Section 4 presents objective results based on numerical comparison, as well as subjective results for visual inspection.

2. MEASUREMENT OF DISPLAY PROPERTIES

A multi-view autostereoscopic display is a nonlinear system that transforms, depending on the observation angle, a digital image into several, somehow distinct, images or views. As described in the previous section, this is achieved by the display architecture i.e. LCD matrix combined with an optical filter. An observer staying within one view (looking at the display from a particular direction) will see only a fraction of the original image, that is, he/she will see a downsampled version of the image. In order to represent the image correctly and foremost to avoid alias errors caused by downsampling it is necessary to filter the image with an antialiasing filter before it is sent to the display. The design of the antialiasing filter requires the knowledge about the frequency properties of the display. A detailed procedure for measuring various properties of autostereoscopic displays is given in [9]. For convenience of the reader, some material from [9] is given in this section with the emphasis put on the frequency properties of the display.

2.1 Sub-pixel visibility versus observation angle

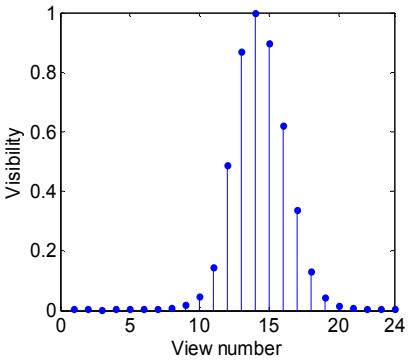
The optical filter of a multi-view display is directionally selective. The apparent brightness of each pixel depends on its intensity, its position in respect to the optical filter, and the position of the observer (distance and observation angle) in respect to the display. The brightness of a sub-pixel can be measured by photographing the display at various angles. In this work, the effect of the optical filter is modelled as *visibility* of each sub-pixel. Visibility of a sub-pixel is the ratio between the assigned intensity and the measured brightness of that pixel. For gamma corrected and normalized images, visibility is a scaling coefficient between 0 and 1. It is assumed, that all sub-pixels, which belong to one view are equally affected by the optical filter, and have maximum visibility for the optimal observation spot of that view. The visibility of each sub-pixel as a function of the observation angle is measured by the following general measurement methodology of five steps. As a case study, this paper presents measurement results for 23" 3D Display AD built by X3D-Technologies GmbH, which is hereafter referred to as *X3D display*. Further details about the measurements can be found in [8] and [9].

The first step is to derive the size of interleaving pattern by observing the behavior of various test patterns. Such test pattern is an image where every *n*-th sub-pixel in a row and every *m*-th sub-pixel in a column are fully lit, and the rest are black. The test pattern with the correct size is fully invisible for most observation angles. The interleaving pattern of X3D display is found to be 8 sub-pixels wide and 12 rows high. The next step is to prepare a group of test images,

where only one sub-pixel per pattern with that size is lit, and to find the optimal observation spot of each group. The sub-pixels which have the same optimal observation spot belong to a single view. The views are numbered by order of appearance of their optimal observation spot. For X3D display there are 24 such groups, which results in an interleaving pattern as shown in Figure 3(a). Finally, test images are generated where only sub-pixels belonging to one view are fully lit. The brightness of each test image is measured from each optimal observation spot. The mean brightness measured on an area of the screen gives the visibility of the sub-pixels, which belong to the same view, as seen from the chosen observation spot. For X3D display, the brightness of each view as seen in front of the centre of the display is given in Figure 3(b). The measurements for other observation spots produce similar curves, with peaks (maximum visibility) shifted to the corresponding view.

	Column 1			Column 2			Column 3			Column 4			Column 5			
	R	G	B	R	G	B	R	G	B	R	G	B	R	G	B	R
ROW 1	2	5	8	11	14	17	20	23	2	5	8	11	14	17	20	23
ROW 2	4	7	10	13	16	19	22	1	4	7	10	13	16	19	22	1
ROW 3	6	9	12	15	18	21	24	3	6	9	12	15	18	21	24	3
ROW 4	8	11	14	17	20	23	2	5	8	11	14	17	20	23	2	5
ROW 5	10	13	16	19	22	1	4	7	10	13	16	19	22	1	4	7
ROW 6	12	15	18	21	24	3	6	9	12	15	18	21	24	3	6	9
ROW 7	14	17	20	23	2	5	8	11	14	17	20	23	2	5	8	11
ROW 8	16	19	22	1	4	7	10	13	16	19	22	1	4	7	10	13
ROW 9	18	21	24	3	6	9	12	15	18	21	24	3	6	9	12	15
ROW 10	20	23	2	5	8	11	14	17	20	23	2	5	8	11	14	17
ROW 11	22	1	4	7	10	13	16	19	22	1	4	7	10	13	16	19
ROW 12	24	3	6	9	12	15	18	21	24	3	6	9	12	15	18	21

(a)



(b)

Figure 3, Measurement results a) derived interleaving pattern of X3D display, b) sub-pixel visibility versus view number, from the optimal observation spot of view 14

2.2 Frequency response of multi-view displays

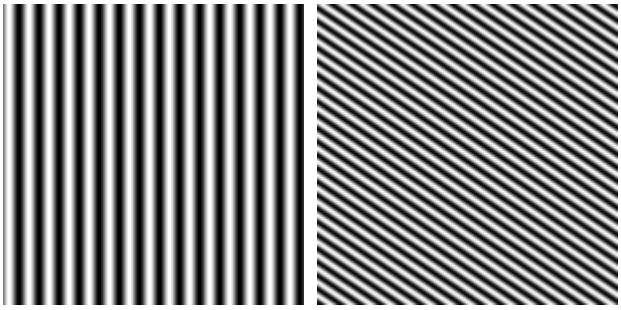
An elegant way for deriving the frequency response of the display, especially when not all display specifications are available, is by using measurements. The main idea in this approach is to generate a set of test images containing signals with various known frequencies, visualize them on the display, and then analyze the output of the display. This procedure is described in the following three subsections. Whereas the approach is illustrated for the X3D display, it is perfectly applicable for any other multiview displays.

2.2.1 Test images

For the purpose of measuring the frequency response of the display, numerous test images (hereafter referred to as input images) have been generated. Each of them is a pattern of a known frequency. Two of those images for frequencies[1] $(f_x, f_y) = (0.2, 0)$ and $(f_x, f_y) = (0.2, -0.3)$ are shown in Figure 4 with the corresponding spectra shown in Figure 5. Here, f_x and f_y refer to frequencies along the x and y axis, respectively. The idea behind generating these images lies in the fact that their frequency behaviour is well known, that is, they have well defined, known, distinct frequency components as

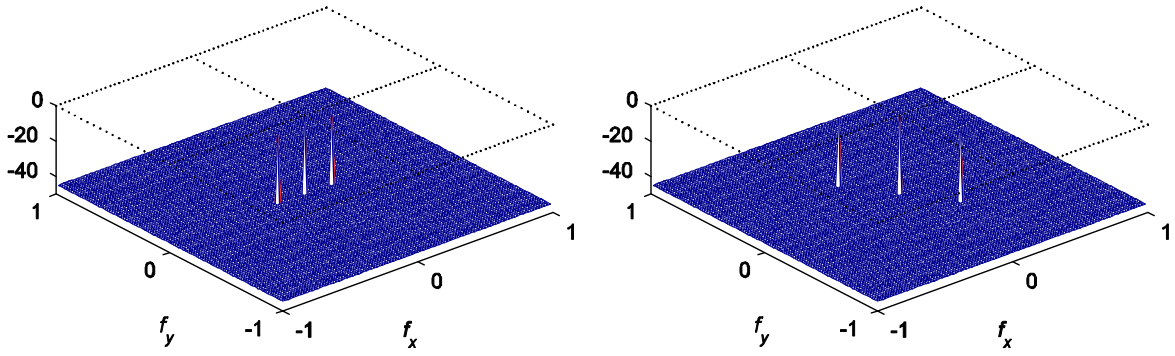
[1] In this paper all frequencies are normalized to $f_s/2=1$ with f_s being the sampling frequency.

it can be seen in Figure 5. In both cases, in Figure 5, the central peak at $(f_x, f_y) = (0, 0)$ is the DC component and should be ignored.



(a) (b)

Figure 4, Example of input (test) images. (a) $(f_x, f_y) = (0.2, 0)$. (b) $(f_x, f_y) = (0.2, -0.3)$.



(a) (b)

Figure 5, Example of input images – Spectra. (a) $(f_x, f_y) = (0.2, 0)$. (b) $(f_x, f_y) = (0.2, -0.3)$.

Several hundreds of those images were generated for sets of frequencies f_x and f_y belonging to the intervals $f_x \in [0, 1]$ and $f_y \in [-1, 1]$ (the signals for $f_x \in [-1, 0)$ and $f_y \in [-1, 1]$ can be easily reconstructed by taking into account symmetry properties of spectra of real signals). By selecting, on the above intervals, a dense grid of frequencies (e.g. $\Delta_f \geq 0.01$), all possible combinations for input images are taken into consideration. It should be pointed out that a denser grid results in a more precise estimation (better resolution) of the frequency response but it also considerably increases the number of required measurements. Therefore, a proper compromise between the required resolution and the number of measurements has to be made. In this paper the step $\Delta_f = 0.025$ has been used. This has turned out to be a good choice for designing antialiasing filters for the display under consideration.

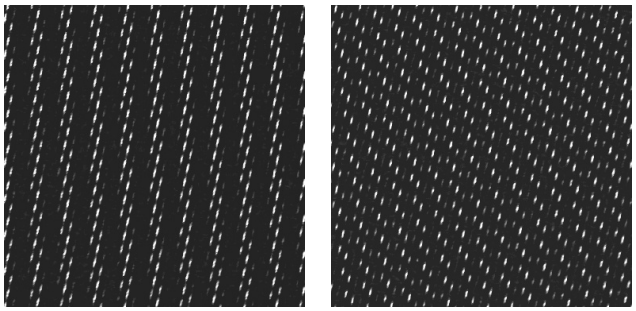
2.2.2 Measurements

The input images, described in the previous section, have been visualized on the display and, by using a high resolution digital camera, photos of the screen have been taken (hereafter referred to as output images). As an example, for the

input images shown in Figure 4 the output images are given in Figure 6 (images have been enhanced for clarity). The spectra of these images are shown in Figure 7.

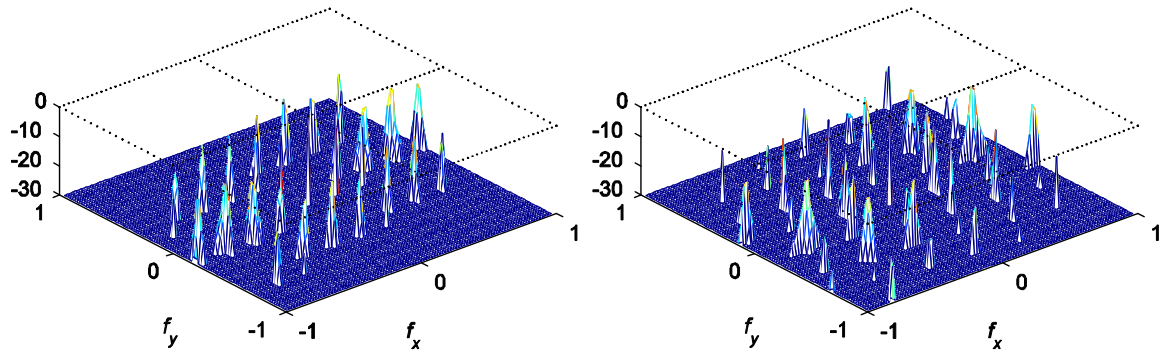
Three observations can be made based on these measurements. First, although each of the input images contains only a single frequency component, the output images contain numerous different frequency components. This is mainly due to the aliasing and imaging effects of the display. As already discussed before, aliasing is the consequence of having multiple views, that is, from one observation angle only part of pixels is visible. This corresponds to downsampling of the original image. Aliasing effects can be removed by a proper antialiasing filter. On the other hand, imaging occurs due to the gaps between visible sub-pixels (See Figure 3(a)). In the spectral domain, imaging can be seen as high frequency components. Unfortunately imaging cannot be avoided or compensated by any filtering as it occurs in the display. Fortunately, as long as no aliasing occurs, it has been experimentally shown that those imaging components are partially suppressed by human visual system (e.g in Figure 6(a) the vertically lines are still seen even if they are heavily broken).

Second, if the frequency of the signal is low, then the signals can be correctly represented on the display (e.g. it is easy to identify vertical lines in Figure 6(a)). However, if the signal frequency is too high then due to the aliasing and imaging effects the image on the display would differ from the original one (e.g. in Figure 6(b), beside the barely visible diagonal lines from Figure 4(b) many other lines are also seen).



(a) (b)

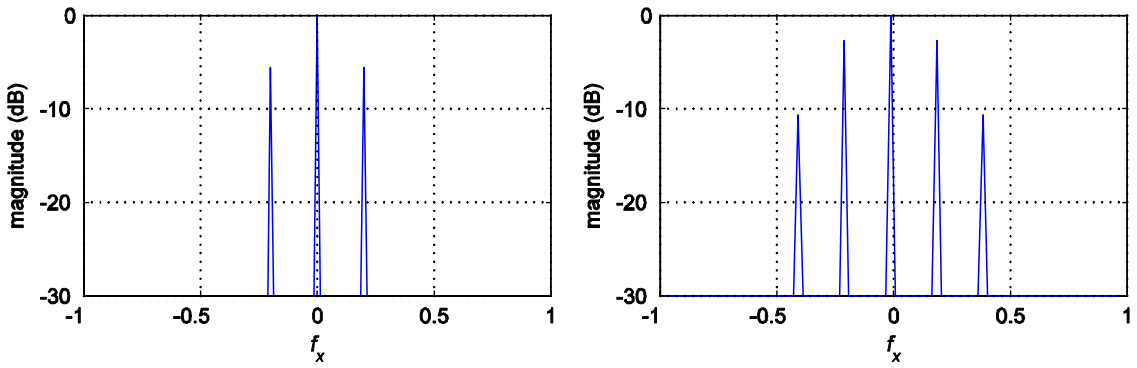
Figure 6, Example of output images (photos taken from the display). (a) $(f_x, f_y) = (0.2, 0)$. (b) $(f_x, f_y) = (0.2, -0.3)$.



(a) (b)

Figure 7, Example of output images – Spectra. (a) $(f_x, f_y) = (0.2, 0)$. (b) $(f_x, f_y) = (0.2, -0.3)$.

Third, the monitor introduces nonlinear distortions as illustrated by Figure 8. Figure 8(a) and Figure 8(b) show the spectra along the x-axis for the input signal $(f_x, f_y) = (0.2, 0)$ and the corresponding output signal, respectively. Although the input signal has only one spectral component (at $f_x = 0.2$), the output signal also contains some higher harmonics (at $f_x = 0.4$) approximately 6-8 dB lower than the main spectral component.



(a) (b)

Figure 8, Nonlinear distortions – Spectra along x-axis for signal $(f_x, f_y) = (0.2, 0)$. (a) Input image. (b) Output image.

2.2.3 Frequency characteristics of the ideal antialiasing filter

In order to analyze the performance of the display in the frequency domain, the spectra of the input and output images derived in the previous two sections were compared. In order to eliminate measurement errors (noise) the spectrum of the output image has been thresholded to -30dB below the strongest frequency component.

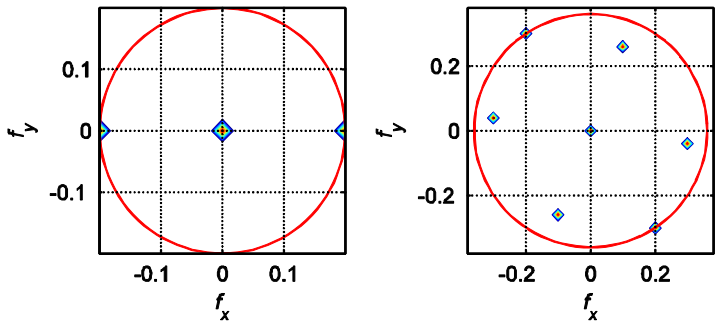
The criteria for determining if a given frequency component passes through the system properly or it is distorted by aliasing and imaging errors was the following: For every input signal of frequency (f_{x_0}, f_{y_0}) it was checked if the contributing aliasing / imaging components contain frequency components that are inside a circle with radius

$$r_0 = \sqrt{f_{x_0}^2 + f_{y_0}^2} ,$$

that is, if they are of a smaller frequency than the original one (in all cases the DC component is ignored). If there are frequency components smaller than the one present in the input signals, this means that the system aliased some of the

frequencies and as a consequence, there will be visible distortions on the display. By removing such frequencies from the input image, aliasing effects can be avoided. Hence, the stopband of the antialiasing filter should suppress all those frequencies.

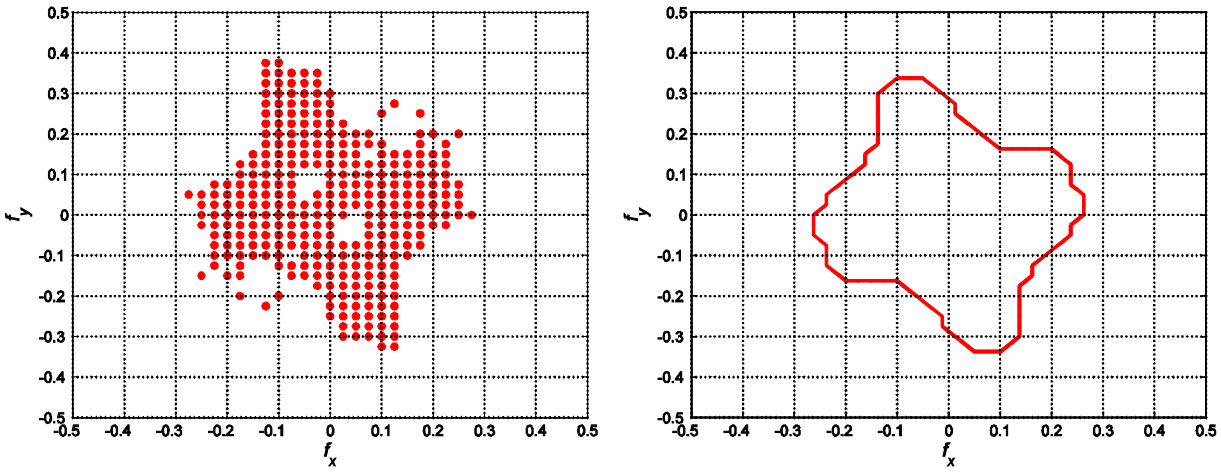
Two examples of spectra (represented as contour plots) of output images are shown in Figure 9. This corresponds to the two examples used in the previous sections. Figure 9(a) and Figure 9(b) are magnified (contour) version of figures Figure 7(a) and Figure 7(b), respectively. In these figures only the part containing frequencies smaller than r_0 (represented by the red circle) is shown. As seen from the figures, in the first example, there are no spectral components that are of a lower frequency than the one used for generating the input image and therefore the image is properly represented on the display. In the second example, the output image considerably differs from the one sent to the display due to the aliasing errors. In the spectral domain this can be noticed by presence of several frequency components that are inside the circle with radius r_0. There is no point in trying to represent this image on the display under consideration, that is, this frequency should be suppressed before visualizing the image on the display.



(a) (b)

Figure 9, Spectra of output images. (a) $(f_{x0}, f_{y0}) = (0.2, 0)$, r_0=0.2. (b) $(f_{x0}, f_{y0}) = (0.2, -0.3)$, r_0=0.36.

By applying the above criteria to all output images, the passband (frequencies that do not cause aliasing) and stopband (frequencies that do cause aliasing) can be classified as given in Figure 10(a). In this figure, the passband is represented by dots. In order to get a smoother filter characteristic that can be used in the filter design, a 5 by 5 median filter has been applied resulting in the desired ideal cut-off frequency of an antialiasing filter shown in Figure 10(b). Such ideal filter would suppress all undesired frequency components in the image resulting in an alias-free image.



(a) (b)

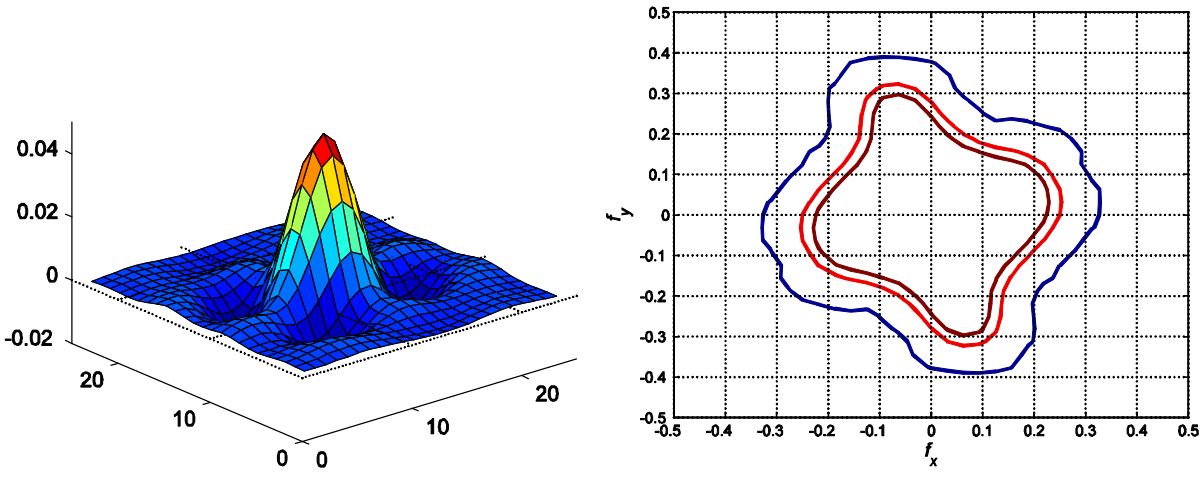
Figure 10, Ideal 2D filter. (a) Passband region estimation based on measurements. (b) Contour of the ideal filter.

3. DESIGN OF ANTIALIASING FILTERS

The discussion in previous sections argued why it is important to filter an image before visualizing it on an autostereoscopic display. In an earlier paper [7], it has been shown that visually good results can be achieved with separable 2D antialiasing filters that were optimized by subjective experiments. However, in practice it is better to have an objective design method that does not depend on subjective testing. Therefore, based on the results of the measurements described in the previous sections, in the following two sections separable and non-separable 2D filters are designed for the display under consideration.

3.1 Non-separable antialiasing filters

For the display under consideration, the shape of an ideal 2D antialiasing filter is shown in Figure 10(b). In this figure, the curve shows the ideal cut-off frequency, that is, the passband of the filter should be inside the contour, and its stopband everywhere else. For designing a non-separable 2D filter approximating this ideal one, the windowing design technique with the Kaiser window of length 24 has been used (e.g. see `fwind2` function in Matlab) [10]. The design results in the 24 by 24 2D non-separable filter with impulse response shown in Figure 11(a). The corresponding magnitude response (contour) of the designed filter is shown in Figure 11(b). The Kaiser window has been selected as a good candidate due to its narrow transition band and flexible attenuation. The variable parameter of the Kaiser window controlling the stopband attenuation has been set to β=2.2. Such selection will ensure a stopband attenuation of at least 30dB that is good enough for the display under consideration.



(a) (b)

Figure 11, 2D non-separable filter. (a) Impulse response. (b) Magnitude response – contour for -3 (innermost line), -6, and -30 dB (outermost line).

The -6dB line in Figure 11(b) approximates the ideal cut-off frequency. Due to the finite transition bandwidth of the designed filter, even after applying it to the input image, some aliasing errors will occur on the display. However the aliased frequencies will be attenuated by the filter (either filter transition band or stopband) and as such they will not be visible. A sharper filter can be generated by increasing the filter order, which in turn, increases the number of multiplication required for filtering the image. On the other hand, filters of a smaller size will approximate the edge of the ideal filter with lower precision. Moreover, sharper filters have also a tendency to cause edge artifacts in filtered images. Therefore, filter size of 24 by 24 has been chosen as a good compromise between the implementation complexity, transition bandwidth, and approximation of the ideal filter.

3.2 Separable filters

The 2D non-separable filter proposed in the previous section is a very good approximation of the ideal one given by Figure 10(b). However, the computational complexity of a 2D filter is rather high. Considerable computational savings are achieved if the 2D filter can be separated into two 1D filters, one filtering in the horizontal direction and one in the vertical direction. As long as similar performances are achieved by separable and non-separable filters, for a similar filter size, the separable filters will be considerably faster.

For deriving a separable 2D antialiasing filter, the 24-view model described in Section 2.1 is utilized. Based on this model, a pattern of visible pixels that can be seen from one observation point has been derived as seen in Figure 3. The Spectrum of this pattern is shown in Figure 12. Due to downsampling/upsampling behaviour of the display, each of the peaks in this spectrum corresponds to a source of aliasing. In order to avoid aliasing, a filter has to be designed in such a way that its passband does not overlap with any of its copies generated by moving its centre to any of those aliasing sources.

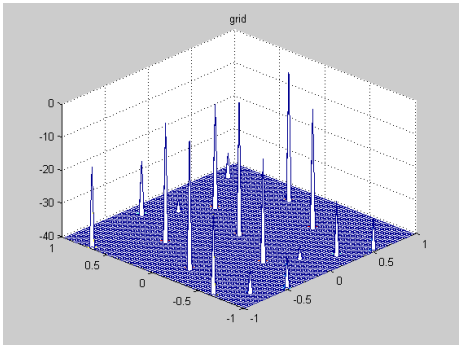


Figure 12, Spectrum of sub-sampling pattern for one view based on the 24-view model.

Additional restriction when using separable filters is that only rectangular-shaped 2D filters can be designed that are symmetrical along the x and y axis. By following basic downsampling principles, it is obvious that there are several different separable filters that can be used as antialiasing filters for this display. This is illustrated in Figure 13. In this figure centres of aliasing terms are marked by red dots with the exact coordinates (frequencies) of each component given in parenthesis as (f_x, f_y) pairs. Moreover only the aliasing terms from Figure 12 pertinent for the design of antialiasing filters are shown. In the figure, three possible ideal filters are drawn (marked as F_1, F_2 and F_3). Each of those filters will perform proper antialiasing, but due to different shapes the visual quality of displayed images will be different. The numerical data for these filters is given in Table I. In the table, f_{cx} and f_{cy} stand for the ideal filter cut-off frequencies in horizontal and vertical direction, respectively.

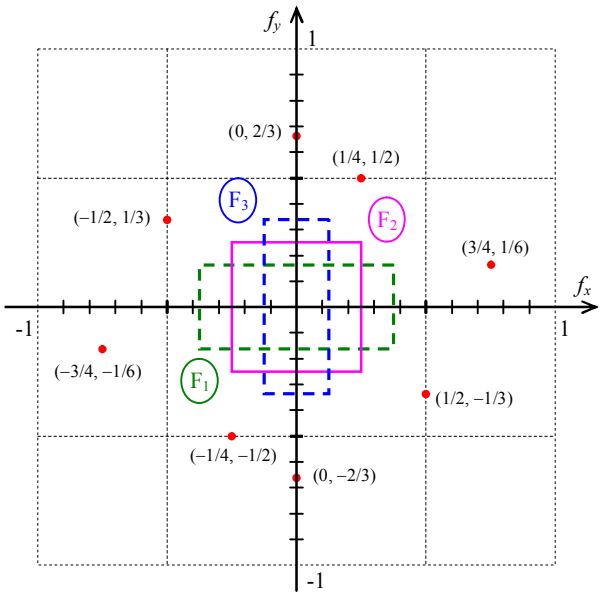


Figure 13, Various ideal separable antialiasing filters.

Table I Horizontal (f_{cx}) and vertical (f_{cy}) cut-off frequencies for ideal separable antialiasing filters

Filter	F_1	F_2	F_3
f_{cx}	3/8	1/4	1/8
f_{cy}	1/6	1/4	1/3

For designing 1D filters with cut-off frequencies given in Table I, as in the case of non-separable design, the windowing technique with the Kaiser window of length 24 has been used. The variable parameter of the Kaiser window has been set to $\beta = 2.2$. The magnitude responses of the designed filters are given in Figure 14. Solid and dashed lines represent the horizontal and vertical filters, respectively. The magnitude responses (contour) of the corresponding 2D filters are given in Figure 15.

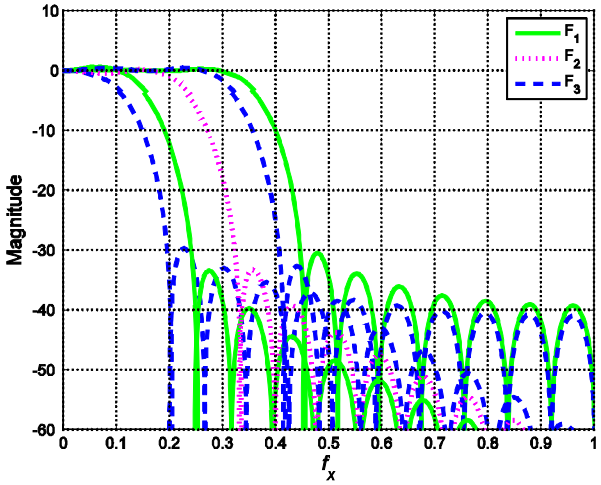


Figure 14, Various ideal, horizontal (solid line) and vertical (dashed line), antialiasing filters of order $N = 23$ for F_1 (green, solid line), F_2 (magenta, dotted line), and F_3 (blue, dashed line). The horizontal and vertical components of F_2 overlap.

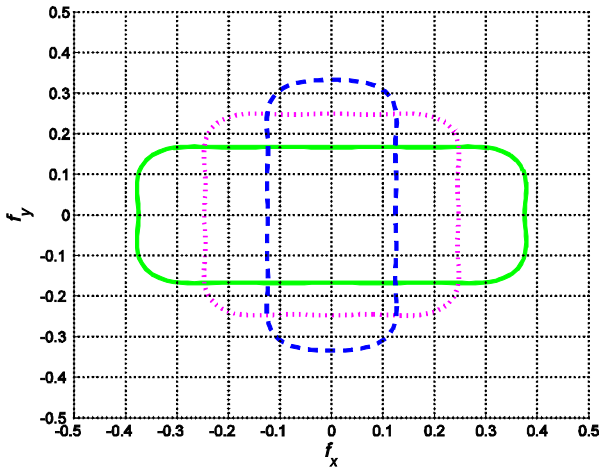


Figure 15, Magnitude responses: – 6dB contour for F_1 (green, solid line), F_2 (magenta, dotted line), and F_3 (blue, dashed line).

4. RESULTS

In this section, the advantage of the proposed methodology with respect to previously suggested antialiasing filter design approaches is demonstrated by objective comparisons of filter performance and computational efficiency and by visual inspection on a set of test images.

4.1 Numerical comparison between filters

The passband size was evaluated as the passband area of the 2D filter or in the case of separable filter, passband area of the corresponding 2D filter, shown in Figure 15. The implementation complexity is given as the number of

multiplications per pixel and is denoted by *C*. The results of both comparisons are given in Table II. Moreover, for completeness of the results, the filters designed in this paper are also compared with the ones presented in [7]. In that work two filters have been suggested, denoted as 'smooth' ad 'sharp' due to their effect on the processed image. The 'smooth' one was aimed at total alias terms suppression while the 'sharp' one was optimized visually to allow for some small amount o aliasing for the sake of better sharpness.

Table II Numerical comparison of various antialiasing filters

	Proposed in this work				Presented in [7]		
Filter	2D	F_1	F_2	F_3	JK	'smooth'	'sharp'
size (length)	24 by 24	24	24	24	48 by 48	15, 18	23, 23
Passband area	.048	0.063	0.063	0.042	0.068	0.033	0.107
C	576	48	48	48	2304	33	46

Several observations can be made based on Table II. First, it is obvious that non-separable filters require considerable higher number of multiplications than the separable ones. Furthermore, the separable filters proposed in this paper are of higher order than the ones denoted as 'smooth' in [7]. This was caused mainly by attempting to widen the passband (e.g. in filter F_2). A wider passband suppress less amount of frequencies in the information part of the signal. However, it imposes also a narrower transition band, and thus a higher filter order, so to ensure effective alias suppression around the passband edge.

Second, when comparing the proposed 2D filters and the one designed by the Jain and Konrad [1], it can be seen that the passband area of the proposed filters is smaller. This is due to the nonlinear distortions (see Section 2.2.2) that exist in the display but are not taken into account by the model used by Jain and Konrad. Moreover, because of these distortions, the filters used in [1] had to be of a higher order to provide shorter transition bandwidths thereby eliminating the alias components caused by the nonlinear distortions.

Third, from the three 1D filters (F_1, F_2 and F_3) proposed in previous section, F_2 has the best approximation of the ideal shape given by Figure 10(b). It will be demonstrated in Section 4.2 that it also performs best in visual inspections.

Forth, when using separable filters some minor aliasing errors are to be expected, because with separable filters it is impossible to get the ideal 2D shape shown in Figure 10(b). Nevertheless, it can be claimed, based on numerous experiments, that this aliasing is tolerable and does not compromise the image quality.

Fifth, the complexity evaluation in the table assumes direct implementation of every filter because in this way it is easy to have a relative comparison between filters. If in the implementation the coefficient symmetry is utilized and/or some other algorithms for fast implementation of a filter are used then the implementation complexity, for all filters in the table, can be further reduced.

4.2 Visual inspection

The performance of different antialiasing filtering is illustrated by presenting the filtering effect on three test images. Each image has been filtered with the set of filters, visualized on X3D display and then photographed.

The first image, denoted as 'Patterns' contains straight lines and patterns with high contrast and varying spatial frequencies. As a use case, it represents 2D geometric content found in a graphical user interface and it is particularly suitable for demonstrating aliasing effects. For example, the slanted lines in the image are at angles, which are most affected by the optical filter of the X3D display. The original test image is presented in Figure 16(a). The same image photographed as visualized on X3D display is shown in Figure 16(b). Structural and colour artefacts due to aliasing are clearly visible. Photographs of the test image, pre-processed with two different filters are given in Figure 16 (c) and (d). Figure 16(c) shows the image as pre-filtered by 2D non separable filter. Figure 16 (d) shows the image as pre-filtered by separable filter F_2. One can see that the image filtered by the non-separable filter exhibits no aliasing artefacts, while the image filtered by F_2 preserves more details, but some elements are still aliased.

The second test image consists of 2D text with variable font size, created by Wordle [11]. The original test image is given in Figure 17(a). The images in Figure 17 (b), (c) and (d) are photographs of the test image, pre-filtered by 2D filter, F_1 and F_2 respectively. Visual inspection shows that 2D filter and F_2 produce comparable results in terms of perceptual quality. Even though filters F_1 and F_2 have equal size of the passband area, images filtered by F_2 are easier to read due to the fact that F_2 has the same throughput in horizontal and vertical direction.

Finally, the filters are visually compared using full-colour, natural 2D image. The image is "Lighthouse" from the Kodak Image Database [12]. The results produced by the non-separable and separable (F_2) filters are quite similar for that image. One can conclude that for natural images containing low or no amount of slanted patterns at 'critical' angles (determined by the topology of the optical filter slant) the performance of the F_2 filter is quite competitive to that of the non-separable filter being closest to the ideal antialiasing filter for the given sampling topology.

5. CONCLUSIONS

In this article, we have proposed a methodology for designing antialiasing filters for autostereoscopic displays. Such displays are characterized by additional optical layer (filter) on top of conventional LCD to create different views for different directions. The attempt to create a number of different views reduces the spatial resolution and in order to cope with this problem, the optical layer is mounted in a slanted manner. While this design offers a good compromise between spatial resolution and number of views, it also specifies a more complex non-rectangular sub-sampling pattern on a sub-pixel level. Our methodology is based on simple, yet precise enough measurements of the aliasing effects on the display. Only basic knowledge about the display, e.g. resolution is required to design and conduct the measurements. This is in contrast with previous approaches proposed in the literature which require more detailed

knowledge about the construction of the displays. Other previous approaches suggested modelling (and sometimes simplifying) the view sub-sampling topology. Such models however cannot predict possible nonlinearity in the optical system which is also source of aliasing artefacts. However, it is well revealed by our measurement methodology and subsequently taken into account in the filter design.

The form of the measured passband is such that requires a 2D non-separable filter with certain order. While designing such a filter, alternatively we have designed also a separable approximation which competed favourably in terms of visual performance and computational cost. The design is fairly automatic and little interaction with the user is required (e.g. selection of filter orders). The design is robust to variations of the input parameters so even a not so precise choice by the user will lead to satisfactory results.

We have considered the case where images are placed at the screen surface (images for the left and right eye having zero disparity). Our future work will consider the case of changing disparity and its influence on the antialiasing filtering.

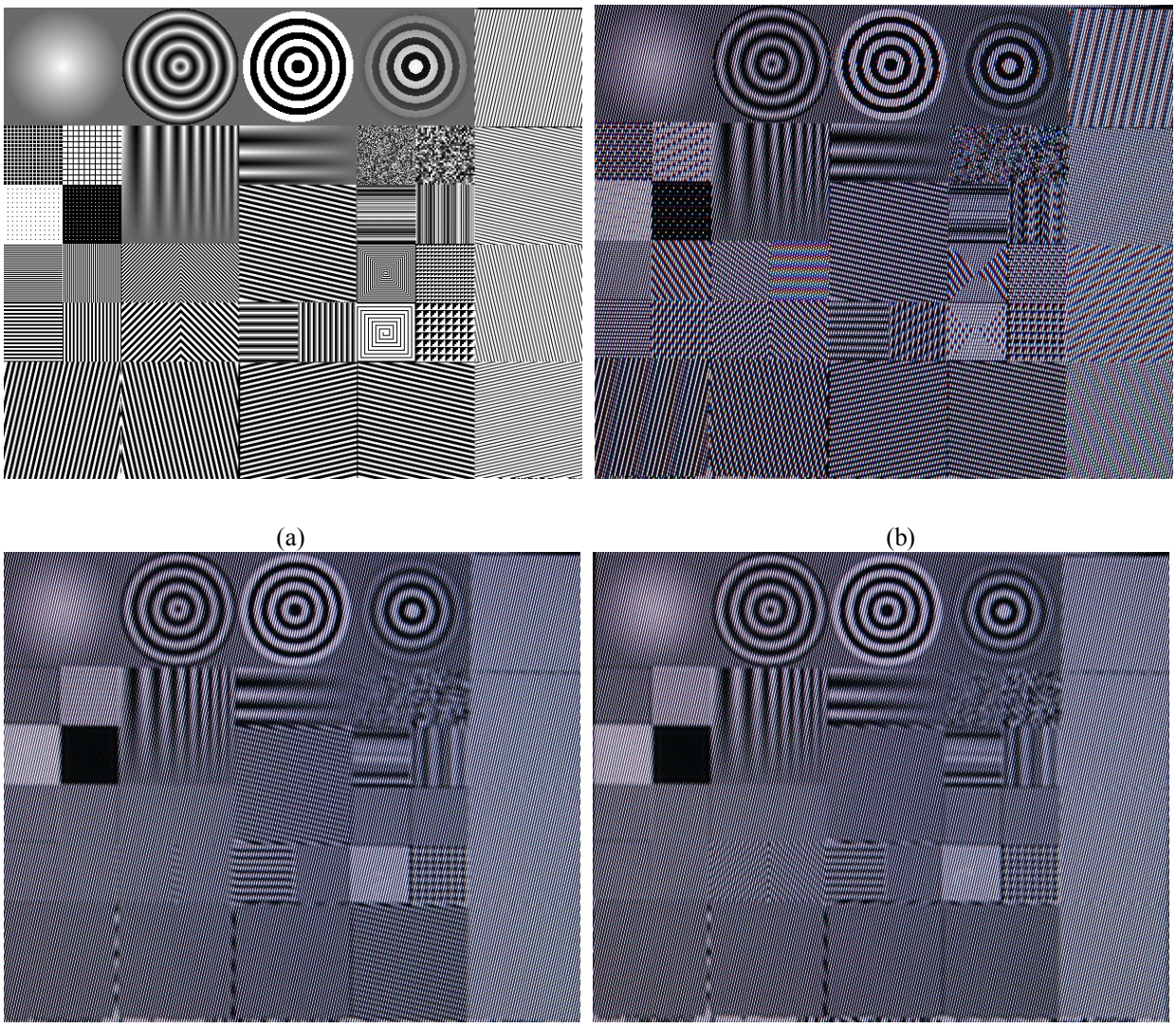


Figure 16, Test image with geometric lines: a) original image, (b-d) images, photographed on X3D display, as follows – b) unprocessed image, exhibiting aliasing, c) filtered with 2D filter, d), filtered with F_2.

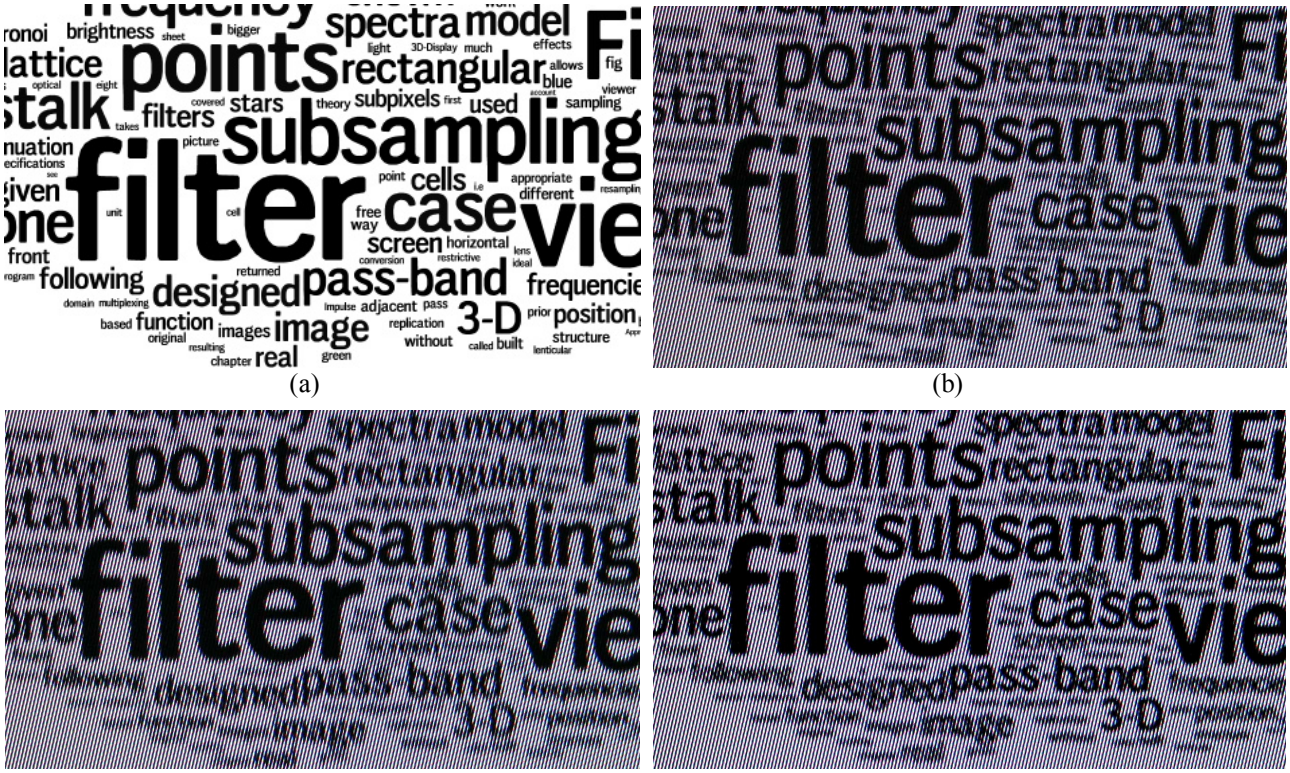
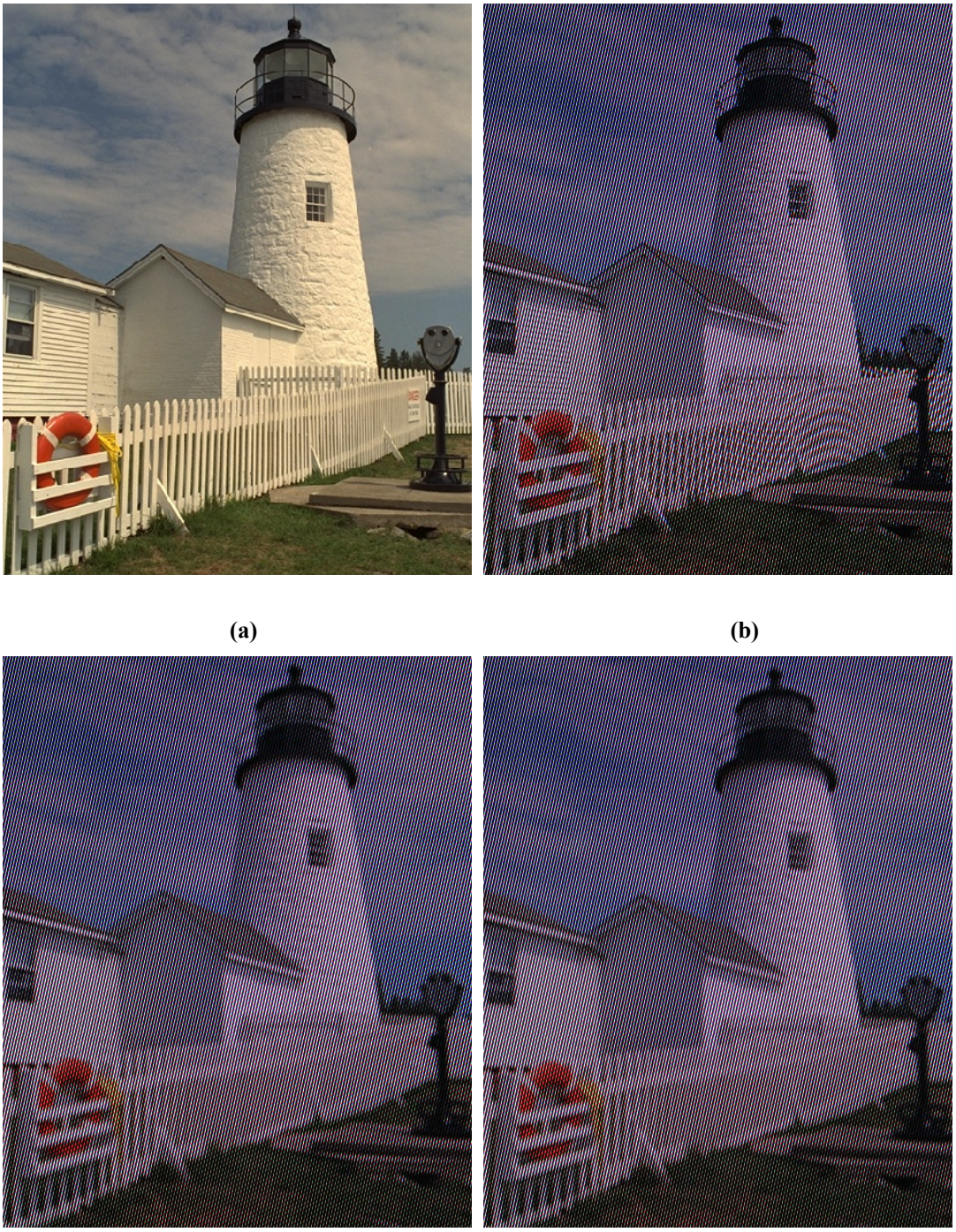


Figure 17, Test image with text: a) original image, (b-d) images, pre-filtered and photographed on X3D display, as follows – b) pre-filtered with 2D filter, c) pre-filtered with F_1, d) pre-filtered with F_2. Note: unprocessed image with aliased text is clearly with the worst quality, but is not shown here because of space limitations.



(a) **(b)**

(c) **(d)**

Figure 18, 2D full-colour test image: a) original image, (b-d) images, photographed on X3D display, as follows – b) unprocessed image, exhibiting crosstalk, c) pre-filtered with 2D filter, d) pre-filtered with F_2.

REFERENCES

[1] C. Van Berkel and J. Clarke, “Characterisation and optimisation of 3D-LCD module design”, in Proc. SPIE Vol. 2653, Stereoscopic Displays and Virtual Reality Systems IV, (Fisher, Merritt, Bolas, edts.), pp. 179-186, May 1997

[2] S. Pastoor, “3D displays”, in (Schreer, Kauff, Sikora, edts.) 3D Video Communication, Wiley, 2005.

[3] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in Proc. SPIE Photonics West 2002: Electronic Imaging, vol. 4660, pp. 212-221, 2002

[4] Zwicker, M., Matusik, W., Durand, F., Pfister, H., and Forlines, C. 2006. Antialiasing for automultiscopic 3D displays. In *ACM SIGGRAPH 2006 Sketches* (Boston, Massachusetts, July 30 - August 03, 2006). SIGGRAPH '06. ACM, New York, NY, 107.

[5] A. Jain and J. Konrad, “Crosstalk on automultiscopic 3-D displays: Blessing in disguise?,” in Proc IS&T/SPIE Symposium on Electronic Imaging, Stereoscopic Displays and Applications, San Jose, CA, Vol. 6490, pp. 649012 (2007).

[6] Moller, C. N. and Travis, A. R. 2005. Correcting Interperspective Aliasing in Autostereoscopic Displays. *IEEE Transactions on Visualization and Computer Graphics* 11, 2 (Mar. 2005), 228-236.

[7] A. Boev, R. Bregovic, A. Gotchev, K. Egiazarian, Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays, in Proc. of *The 2009 International Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2009*, Helsinki, Finland, 2009

[8] A. Boev, A. Gotchev and K. Egiazarian, “Crosstalk measurement methodology for auto-stereoscopic screens”, Proc. of 3DTV Con, Kos, Greece, 2007

[9] A. Boev, R. Bregovic, A. Gotchev, “Measuring and modeling per-element angular visibility in multiview displays”, *Special issue on 3D displays, Journal of Society for Information Display*, to be published

[10] S. K. Mitra, Digital signal processing: A computer based approach, New York: McGrow-Hill, 2005

[11] Wordle, software for generating “word clouds”, available online at http://www.wordle.net

[12] Kodak image database, available online at ftp://ftp.kodak.com/www/images/pcd

[P04] A. Boev, R. Bregovic, A. Gotchev, “Measuring and modeling per-element angular visibility in multiview displays”, *Special issue on 3D displays, Journal of Society for Information Display,* Sept. 2010 Vol. 18, No. 09, pp. 686–697

This page left blank

Measuring and modeling per-element angular visibility in multi-view displays

Atanas Boev
Robert Bregovic
Atanas Gotchev

Abstract — Multi-view displays employ an optical layer which distributes the light of an underlying TFT-LCD panel in different directions. Certain properties of the layer create specific artifacts, such as ghost images, moiré patterns, and masking. The layer was modeled as an image-processing channel, and the display parameters related with the model were identified, which are important for the design of image-processing algorithms for artifact mitigation. The identified parameters are interleaving pattern, angular visibility, and frequency throughput of the display. A methodology for deriving these parameters for an arbitrary LCD-based multi-view display are presented, which does not require precisely positioned measurement equipment. As a case study, measurement and modeling results for a particular multi-view display are also presented.

Keywords — *Multi-view displays, angular visibility, cross-talk measurement, optical measurements, 3-D displays, visual quality.*

DOI # 10.1889/JSID18.9.686

1 Introduction

Multi-view displays are a class of autostereoscopic displays, which can be used without the need of special glasses and can be watched by multiple users simultaneously.[1–7] Multi-view displays generate multiple observations of a scene, each one seen from a different angle. Usually, the image is formed on a TFT-LCD. An additional directionally selective *optical layer* mounted on top of the LCD redirects the light of the subpixels in different directions.[1,4,7,18] The layer is either a *parallax barrier*, which blocks the light in some directions[7] or *lenticular sheet*, which works by refracting the light.[8] The apparent brightness of each subpixel is a function of the angle. The group of subpixels which is visible from one direction forms an image, known as *view*.[1,7,9] From a certain spot in front of the display, all subpixels that belong to a view are seen with maximal brightness. Such a spot is referred to as an *optimal observation spot* for the corresponding view. Outside of that spot, there is a larger *view visibility zone*, in which the view is still visible, albeit with diminished brightness. In LCD-based multi-view displays, views are spatially multiplexed.[1,2,9] The process of mapping multiple images to the views of one display is called *interdigitation*,[9] *view multiplexing*,[1] *interlacing*,[3,5] or *view interleaving*.[14] The latter term is adopted in this paper. The relation between the position of a subpixel and the view it belongs to is given by an *interleaving map*. Since both the LCD and the optical layer have a repetitive structure, the interleaving map can be described by a periodic *interleaving pattern*.[7,9] The pattern is spatially independent – the angular visibility of a subpixel depends on its position in respect to the pattern, but not on its absolute position in respect to the display.

The design of a multi-view display is a trade-off between observation convenience and visual quality. The added convenience in using multi-view display comes at the expense of limited brightness, contrast, and resolution.[1,9,10] The optical layer is a source of specific visual artifacts.[1] These are moiré patterns caused by aliasing,[9,11,13] ghost images caused by cross-talk,[7,10,19] and masking artifacts caused by the optical layer behaving like an up-sampling block. The masking artifacts manifest themselves as a fine mesh superimposed over the image, an effect that can be regarded as imaging, as discussed in Sec. 5. The influence of these artifacts on the visual quality of a 3-D scene depends both on image content and optical parameters of the display. It is possible to use image-processing methods to mitigate these artifacts[5,9,13,14] and the effectiveness of the image processing methods depends highly on the information about the optical characteristics of the display, which is, however, rarely available to the end user.

There are various methods for assessing the optical quality of the display; for example, using directional scanning for 2-D[15] and 3-D[10] displays. In Ref. 8 the authors propose an extensive list of optical parameters which can be measured for characterization of autostereoscopic 3-D displays. In this paper, we aim to identify and measure the parameters that can be used for visual optimization. Thus, we are interested in the parameters that are important from the image-processing point of view, rather than in ones which describe the optical quality of a 3-D display.

The paper is organized as follows. In Sec. 2, we propose a model, which considers a multi-view display as an image-processing channel. The model is used to describe the typical visual artifacts in signal-processing terms. The parameters in this model identify which characteristics of the display are needed in visual optimization algorithms. In Secs. 3 and 4, we present a simple, yet effective, methodology to measure and model these parameters. More specifically, in Sec. 3, we describe a way to derive the interleaving topology, and in Sec. 4 we propose a method for finding the

The authors are with Tampere University of Technology, Department of Signal Processing, P.O. 553, Tampere, TRE 33101, Finland; telephone +358-3-3115-4959, e-mail: atanas.boev@tut.fi.

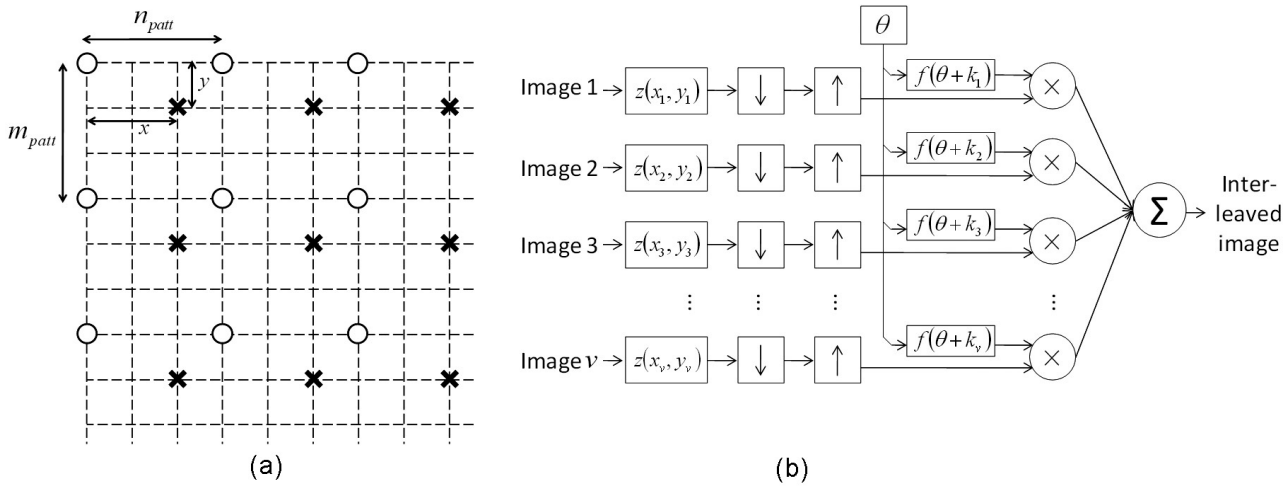


FIGURE 1 — Model of the optical layer effect as an image processing channel: (a) interleaving map as a set of nonoverlapping lattices and (b) interleaved image as a weighted sum of sampled images.

angular visibility of each display element by combining measurement data from several points. Finally, in Sec. 5, we characterize the properties of the display in the frequency domain. As a case study, this paper presents measurement results for 23-in. 3-D display AD built by X3-D-Technologies GmbH, which is hereafter referred to as *X3-D display*.

2 Multi-view display as image-processing channel

Multi-view 3-D displays aim to generate multiple images, each one seen from a different observation angle. The optical layer mounted above the screen surface acts as a directionally selective filter and applies an angular luminance function to each subpixel of the display. The angle, at which the angular luminance has its peak value, determines the optimal observation direction of the subpixel. There are groups of subpixels with similar angular luminance functions, which are simultaneously visible, thus creating the illusion of an image which is visible from certain angles and invisible from others. The number of groups with a similar angular luminance function determines the number of views generated by the display.

The interleaving map can be represented as a set of non-overlapping lattices, where each lattice contains subpixels from a single view only.[9] On an image with the full resolution of the LCD, each of these lattices acts as a rectangular subsampling pattern with a different offset. An example is shown in Fig. 1(a), where the intersecting dotted lines mark the position of LCD subpixels; one lattice is marked with circles and another is marked with crosses. The horizontal step of the lattice is equal to the width of the interleaving pattern [denoted with n_{patt} in Fig. 1(a)] and the vertical step equals to the height of the pattern (denoted with m_{opt} in the same figure). All the subpixels that appear in column $k_1 * n_{patt}$ and row $k_2 * m_{patt}$, where k_1 and k_2 are integers, belong to a group with equal angular visibility. The subpixels that appear in column $k_1 * n_{patt} + x$ (for $0 \le x < n_{patt}$) and row $k_2 * m_{patt} + y$ (for $0 \le y < m_{patt}$) also belong to a group with equal angular visibility. It is possible that more than one of these groups belong to the same view.

A model of a multi-view display as an image-processing chain is shown in Fig. 1(b). We assume that as input we have v images with full resolution, which have to be mapped to v views generated by the display. Out of each image, only subpixels that belong to the corresponding view are used. This is modeled by a 2-D downsampling operation. Since the views are spatially multiplexed, each image gets sampled with different offset, represented by image shift $z(x, y)$, where x, y are the horizontal and vertical offset. On the display, the subsampled image is represented in its original size. The visible subpixels appear either surrounded by black stripes by the parallax barrier or enlarged by the lenticular sheet. This effect is modeled as an upsampling stage, where the introduced samples are either set to zero, or are a repetition of the same sample value. Since several groups of p subpixels can belong to the same view, downsampling and upsampling in the most general cases, is not performed on a rectangular grid.

The angular luminance is modeled as a function of the observation angle. We model the effect of the optical layer on the brightness of underlying pixels as *visibility* – the ratio between the relative brightness of a view and the maximum brightness of the display as seen from the same angle. The function $f(\theta + k_v)$ gives the visibility of view v from observation angle θ. We refer to such function as *horizontal angular visibility*. We assume that the function is the same for all views, with the peak visibility of each view occurring at different observation angle. With k_v we denote the offset of the function for view v.

Using the proposed model, typical visual artifacts on multi-view displays can be explained from the signal-processing point of view. Aliasing occurs if the source images have not been suitably pre-filtered with an anti-aliasing filter before downsampling. The design of an anti-aliasing filter relies on knowledge of the interleaving topology.[9,13,14]

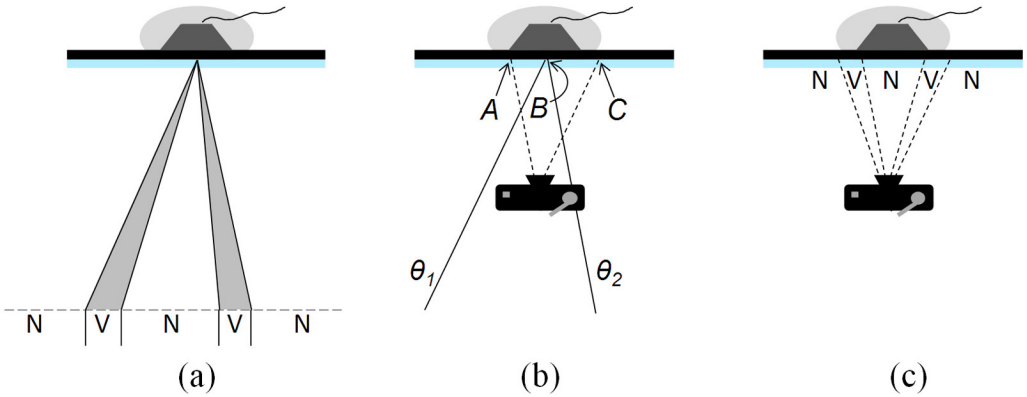


FIGURE 2 — Angular visibility of multi-view display: (a) visibility separation, (b) observation angles when taking a close shot, and (c) close observation of angular visibility.

The topology can be derived by finding n_{patt}, m_{patt}, and all combinations of x and y that belong to the same view. Ghost images occur when images with different offset are simultaneously visible – this effect can also be regarded as cross-talk. Cross-talk mitigation algorithms need knowledge of the angular visibility function,[13,19] which is denoted as $f(\theta + k_v)$ in our model. Due to the optical layer, the visible parts of subpixels have a non-rectangular shape[16–18] and the gaps between them are directionally oriented. The presence of gaps creates effects similar to the ones caused by upsampling in the absence of a post-filter. In sampling and interpolation literature, the effect is denoted as "imaging" and the filters tackling it are known as anti-imaging filters. In the case of multi-view displays, this effect is best quantified by analyzing the performance of the display in the frequency domain.

3 Deriving the interleaving pattern

In some cases, the description of the interleaving pattern is partially or fully missing in the documentation provided along with the display. This prompts proposing a methodology for deriving the pattern. Our proposed methodology has four steps – finding the minimal width and height of the pattern (n_{patt} and n_{patt} in Fig. 1), obtaining the number of views generated by the display, and finally, deriving the complete interleaving pattern.

Ideally, a view should be visible with maximum brightness from a limited range of observation angles and be invisible from anywhere else. An example of angular visibility of a view is given in Fig. 2(a), with "visible" and "invisible" regions marked with "V" and "N," respectively. We refer to the ratio between the width of "N" and "V" regions as *visibility separation*. A group of subpixels with similar angular visibility has a higher overall N/V ratio than a group where each subpixel has a different optimal observation point. The "optimal" interleaving pattern of a 3-D display is one that separates the subpixels into groups that yield the highest visibility separation.

In order to study the angular visibility of a subpixel, one can selectively activate groups of subpixels and perform an angular scan,[5] or use Fourier optics to study a point on the screen from many angles simultaneously.[20] Our approach is to observe the display from a distance shorter than the optimal one, utilizing the space invariance of the pattern. In this approach, the visibility of multiple subpixels as seen in one camera position is related to the visibility of one subpixel as seen from multiple camera positions. As exemplified in Fig. 2(b), point "A" observed from close distance is seen from the same angle as point B observed from the optimal observation distance at angle θ_2. Similarly, point "C" as seen from a closely positioned camera should have the same visibility as point "B" from observation angle θ_2. We refer to images taken from a closely positioned camera as *close shot* images. In the close shot, a horizontal line of the screen is expected to have visibility proportional to the horizontal angular visibility of a point from the optimal observation angle, as seen in Fig. 2(c). The higher the angular visibility ratio [as seen in Fig. 2(a)] is, the higher is the N/V ratio of the horizontal line [as seen in Fig. 2(c)]. In our experiments, the distance between display and camera was 8 cm.

3.1 Finding pattern width

The optimal width of the interleaving pattern is found by observing a row of subpixels, probing for different pattern widths, and selecting the set with the highest visibility separation.

First, one should take a close shot of a square with a known size in subpixels and use it to estimate the camera orientation and to determine the ratio k between display pixels and pixels in the acquired photograph. This allows estimating where each photographed subpixel appears on the photo, regardless of the exact camera placement. The next step is to activate every n-th subpixel in one row [as seen in Fig. 3(a)], and take a close shot. As the line consists of discrete subpixels, the angular visibility of the view would appear sampled at discrete angles. In order to evaluate the visibility separation, one should distinguish dark pixels in the "N" regions masked by the optical layer and dark gaps in "V" regions caused by inactive subpixels. The width of the gaps in the "V" zones is proportional to the currently probed

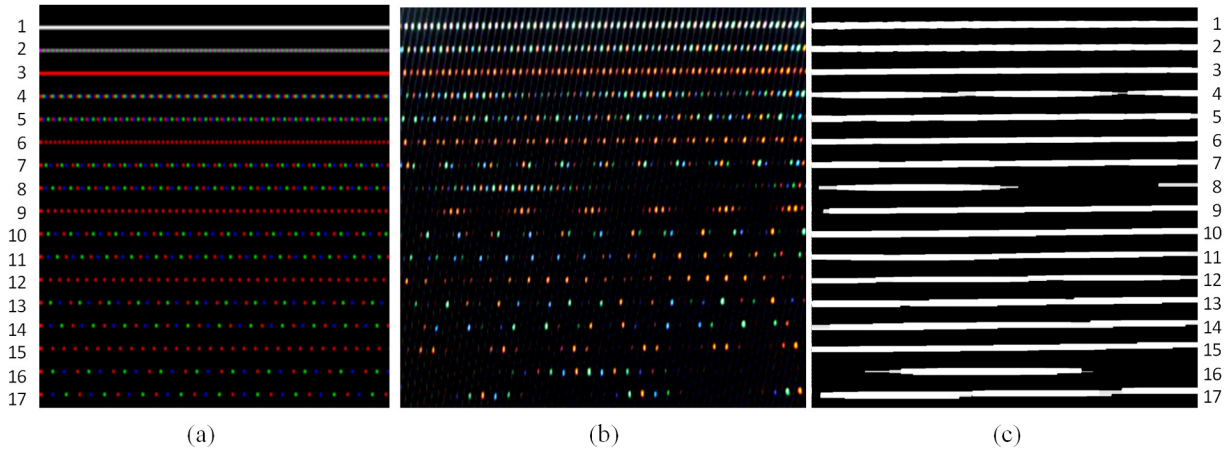


FIGURE 3 — Deriving the width of interleaving patten: (a) test image containing lines with different step size, (b) close observation of the test image, and (c) test image, processed with maximum value filter with coressponding window size.

pattern width n. To make the distinction, we apply a maximum value filter in a horizontal direction using window size of $n * k/3$ (k is divided by three to obtain the display ratio per subpixel).

By probing for various values of n, one can obtain the visibility separation for interleaving pattern with the corresponding widths. The maximum n is bounded by the expected number of views: we used $1 < n < 64$ in our experiments. A test image containing lines with different n can be seen in Fig. 3(a). The same image, as seen on a close shot of the X3-D display, is in Fig 3(b). The processed image, where each line is filtered with the corresponding maximum value filter, is shown in Fig. 3(c). By counting consecutive black pixels in each row, one can find the minimal n which has the highest visibility separation. In our experiments, this is $n_{patt} = 8$.

3.2 Finding pattern height

The optimal height m_{patt} of the interleaving pattern can be found by testing patterns with optimal width n_{patt} and variable height m. The test image for pattern $n \times m$ has subpixels on every n-th subpixel column and every m-th pixel row lit. Each row in that image has optimal visibility separation, but if the optical layer is mounted at a slant, the position of the "V" zones might differ across the rows. For some values of m, the optimal observation directions for each row coincide, making all active subpixels simultaneously visible from certain observation directions. If one considers the mean brightness of all visible subpixels, the optimal m is the value which yields the highest visibility separation for the display as a whole.

Unfortunately, from a close position the angle from which the display rows are seen varies in the vertical direction as well. Thus, from a single close shot it is impossible to predict whether the optimal observation points of each row would coincide at some distance. However, most multi-view displays aim to provide views which spread in the horizontal direction only.[1,7,18] In that case, the angular visibility of the display should change frequently in the horizontal and less often in the vertical direction. In the close shot, this corresponds to vertical lines with minimal slant. In order to distinguish nonvisible from inactive subpixels, one can apply a

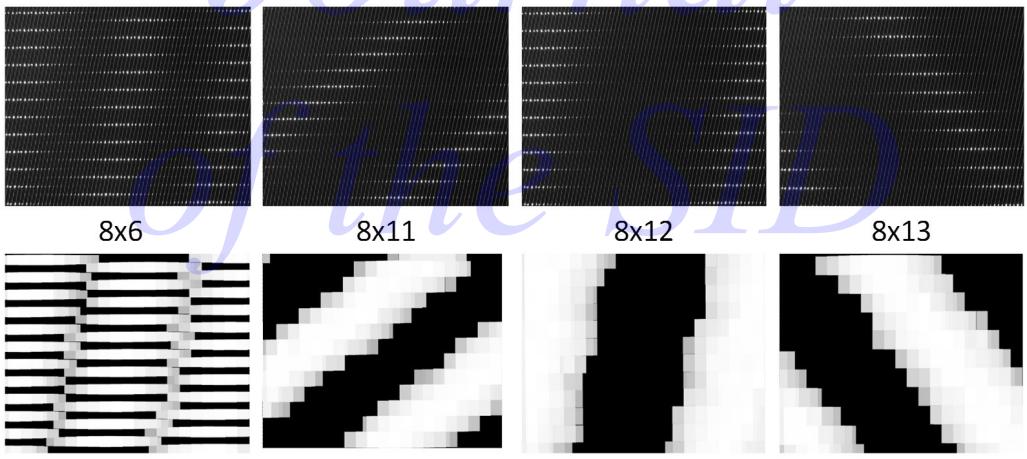


FIGURE 4 — Deriving the height of the interleaving pattern: top row – close observations of test patterns with various heigths; bottom row – the same, processed with maximum value filter with corresponding window size.

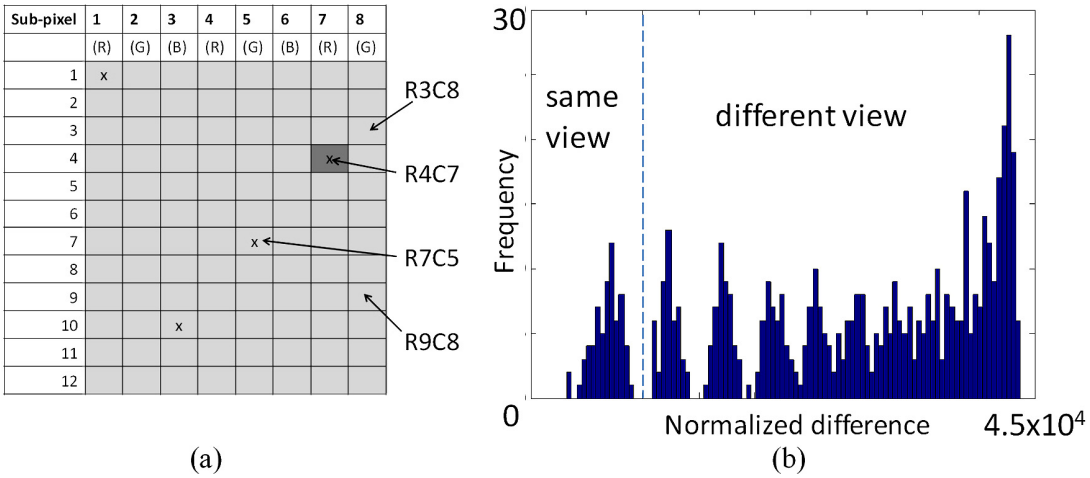


FIGURE 5 — Single-subpixel test patterns: (a) position of subpixel in the pattern (pixels marked with "x" are found to belong to the same view) and (b) normalized differences between close observations of all test patterns.

two-dimensional maximum value filter with window size of $(n * k/3) \times (m * k)$, where n, m are the size of the pattern used in each test. The filtered image with lowest frequency in the vertical direction corresponds to the optimal size of the interleaving pattern.

The top row of Fig. 4 shows close shots of different test patterns on X3-D display, and the bottom row shows the corresponding processed images. In our measurements, we used the position of the biggest peak in the 2-D frequency response to determine the slant of the lines, and found pattern size of 8×12 to be the optimal one.

3.3 Number of views

It is possible that a display with interleaving pattern of $n_{patt} \times m_{patt}$ generates less than $n_{patt} \cdot m_{patt}$ views, *i.e.*, there are subpixels with different coordinates in the pattern which belong to the same view. This case can be tested by building a test image using the optimal pattern size with only one subpixel in the pattern lit, as shown in Fig. 5(a). We refer to these images as *test pattern*, followed by the row and column where the active subpixel is positioned. For example, test pattern "R4C7" is a test image that consists of a periodic pattern with size $n_{patt} \times m_{patt}$, where the subpixel on row 4, column 7 is lit. The total number of such test-pattern images is $n \cdot m$. If two test patterns contain pixels that belong to the same view, they would have identical angular visibility and contribute light to the same regions in the close shot.

The close shots need to be processed with a max value filter with a window size of $(n_{patt} * k/3) \times (m_{patt} * k)$ as is done in the previous section. Test patterns with subpixels which belong to one view should produce slightly different output since subpixels of those patterns appear on different coordinates of the display. However, test patterns with different angular visibility should produce noticeably different results. The norm of difference between two output images can be used to determine if two test patterns belong to the same view or not. One can calculate the l^2-norm of the difference between each pair of filtered images, build a histogram of all norms, and find the threshold between "similar" and "different" norms, as shown in Fig. 5(b). The first group in the histogram represents differences between "similar" patterns that belong to the same view.

Close shots of the X3-D display exhibiting some test patterns and their filtered counterparts are given in Fig. 6. In this example, subpixels "R4C7" and "R7C5" were found to belong to the same view. Test patterns with normalized difference lower than the threshold were grouped together. The whole set of 96 test patterns produced 24 distinctive groups of four patterns each, which means that X3-D display is able to generate 24 distinctive views. Subpixels, which belong to the same group, are marked with "x" in Fig. 5(a).

3.4 Interleaving topology

The final step is to assign a view number to each group of subpixels with similar angular visibility. The order of the views is arbitrary, but for practical reasons it is preferred that views with neighboring observation zones have consecutive numbers. To enumerate the views in that order, one should find the visibility zone of each view.

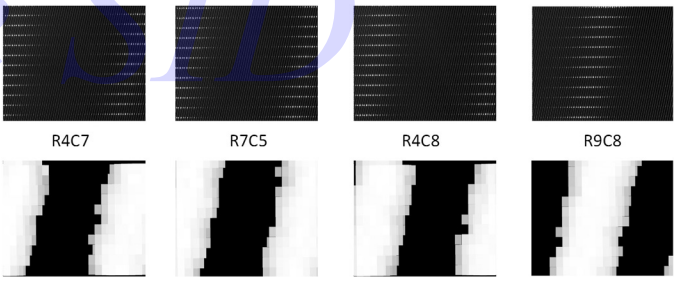
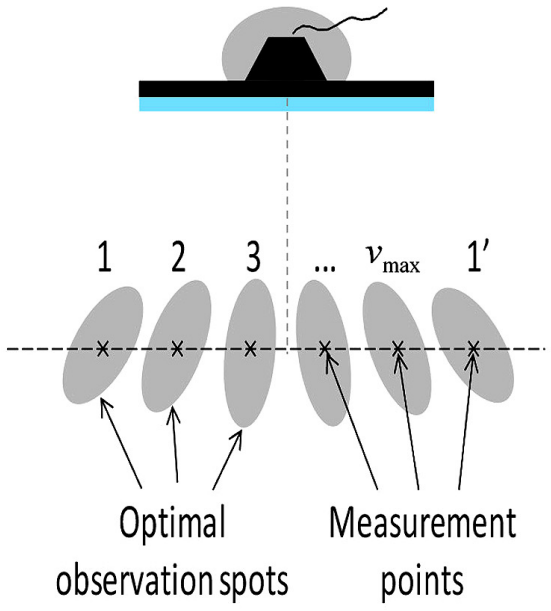


FIGURE 6 — Deriving number of views: top row – close observations of various test patterns; bottom row – the same test patterns after filtering. Subpixels which belong to the same view produce similar close observations [the similarity threshold is shown in Fig. 5(b)].



(a)

Pixel column 1 Pixel column 2 Sub-pixels columns

Pixel rows	1 (R)	2 (G)	3 (B)	4 (R)	5 (G)	6 (B)	7 (R)	8 (G)	9	10	11	12
1	2	5	8	11	14	17	20	23	2	5	8	11
2	4	7	10	13	16	19	22	1	4	7	10	13
3	6	9	12	15	18	21	24	3	6	9	12	15
4	8	11	14	17	20	23	2	5	8	11	14	17
5	10	13	16	19	22	1	4	7	10	13	16	19
6	12	15	18	21	24	3	6	9	12	15	18	21
7	14	17	20	23	2	5	8	11	14	17	20	23
8	16	19	22	1	4	7	10	13	16	19	22	1
9	18	21	24	3	6	9	12	15	18	21	24	3
10	20	23	2	5	8	11	14	17	20	23	2	5
11	22	1	4	7	10	13	16	19	22	1	4	7
12	24	3	6	9	12	15	18	21	24	3	6	9
13	2	5	8	11	14	17	20	23	2	5	8	11
14	4	7	10	13	16	19	22	1	4	7	10	13
15	6	9	12	15	18	21	24	3	6	9	12	15

(b)

FIGURE 7 — Deriving the interleaving topology: (a) measurement points in the approximate centers of the optimal observation spots (view from the top); (b) interleaving pattern for X3-D display.

A set of v_{max} test images is prepared, where v_{max} is the derived number of views. In each test image, all subpixels belonging to the same view are lit. We refer to these images as *single-view* ones. By observing a single-view image on the display, one can search for an observation position from which the display is seen with maximal mean brightness. This position is the optimal observation spot of the corresponding view.

In our experiments, we could identify optimal observation spots with fuzzy borders, appearing approximately on a line, in front of the display, as shown in Fig. 7(a). Because this result is consistent with the measurements of other displays,[5,20] we approximated optimal observation spots to be equidistant on the line, as marked with "X" in Fig. 7(a). In our measurements, the distance between the display and the line was approximately 140 cm.

In our experiments, we enumerate the zones from left to right, so that middle zone numbers are aligned with the center of the display. By labeling the subpixels with corresponding numbers, we obtain the interleaving pattern for X3-D display as shown in Fig. 7(b).

4 Angular visibility of each subpixel

In general, doing an angular scan to obtain the visibility as a function of the angle requires precise positioning – otherwise the angular visibility curve is sampled at irregular intervals. In our method, we measure the visibility of each view at arbitrary points along the optimal observation distance, and search for a single function that gives the best fit for all measurements regardless of the observation point.

4.1 Measurements

As a first step, one needs to prepare all the single-view images that correspond to the views generated by the display. Then the measurement points have to be selected as close as possible to the centers of the visibility zones. In our experiments, 25 measurement points were selected as marked with "X" in Fig. 7(a). Twenty-four of them are in the visibility zones of each view, and the last one is in the visibility zone of the first replica of view 1. Observation point 13 appears in the center.

The next step is camera calibration. Camera sensitivity and aperture should be set to a minimum to minimize CCD noise. Camera exposition should be chosen such that the maximum pixel value in the shots is under the range in which camera response gets into saturation. Then the camera response function should be linearized in the fashion described in Ref. 22. In our experiments, we used 16 gray-level test images where all pixels were set to values between 0 and 255, with step of 16. All the test images were shot in a dark room from measurement point 13.

The last step is measuring the angular visibility. One should prepare v_{max} single-view images, where v_{max} is the number of views generated by the display. Additionally, one *white image* with all subpixels set to maximum brightness and one *black image* with all subpixels set to zero are needed. From each observation position the white, black, and all single-view images should be photographed. The set of images shot from one point is rectified using the corners of the display in the white image. For each single-view image, the mean brightness in the center of the display is measured and normalized to range from 0 to 1, where 0 is the mean

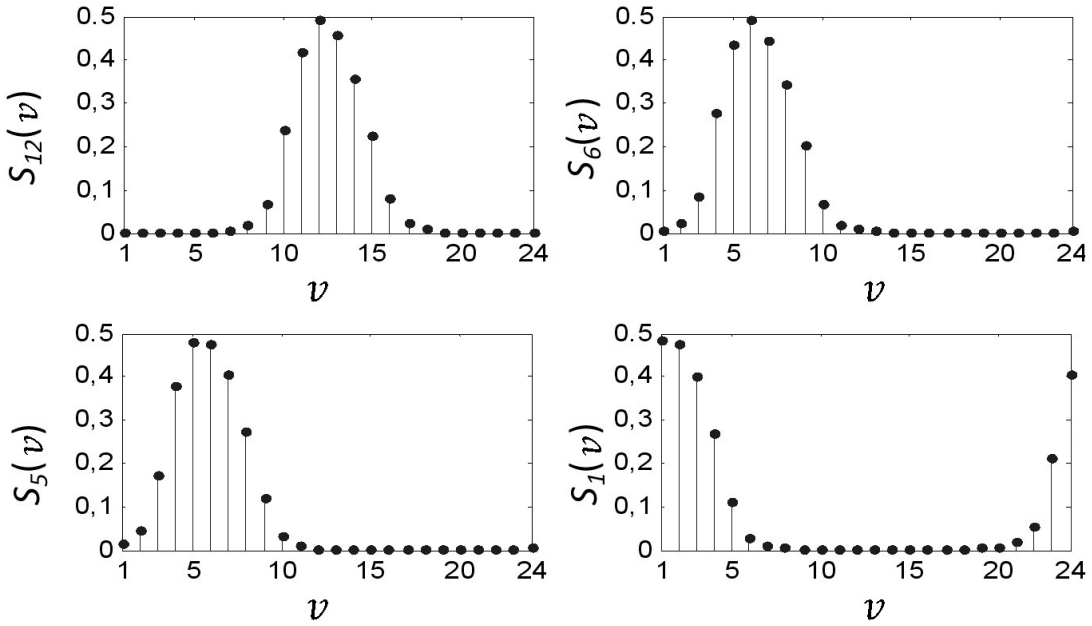


FIGURE 8 — Measurement results for X3-D display for measurement points 12 (top left), 6 (top right), 5 (bottom left) and 1 (bottom right).

brightness measured for the black test image and 1 is the brightness measured for the white test image.

In our measurements, from each measurement point, we took 26 shots of the X3-D display, showing the black, white, and 24 single-view test images. After normalization, we obtained 25 sets of 24 measurements, indicating the brightness of each view relative to the full brightness of the display for each observation position. Because of the normalization, the maximum measured brightness is approximately the same (close to 0.5) for all angles. We used the normalized measurement as visibility $S_p(v)$, where v is the view number and p is the measurement point. The results for four measurement positions are shown in Fig. 8.

4.2 Modeling angular visibility

If the measurement point is displaced from the center of a visibility zone, the visibility function gets sampled with an offset, and the maximum value of that function falls in between two samples. However, judging by the measurement results in other work,[7,20] we assume that the visibility curve for all observation points can be closely approximated by the same function, which has its peak occurring in the optimal observation spot for the corresponding view. We use τ_p to denote the offset of the visibility curve. Integer values of the offset τ_p correspond to the angular visibility in an optimal visibility spot, and noninteger values of τ_p correspond to the angular visibility between spots. Based on this assumption, one can search for a single function that closely fits measurements for all positions regardless of possible offset τ_p. We decided to fit a periodized Gaussian function,

$$G_{a,\sigma,\tau}(v) = \sum_{k=-\infty}^{+\infty} a e^{\frac{(v-\tau-kv_{\max})^2}{2\sigma^2}}, \quad (1)$$

where $v = 1, \ldots, v_{\max}$ and search for optimal (a, σ) that will give the minimum fit error in

$$\arg\min_{a,\sigma\in R^+}\left(\sum_{p=1}^{p_{\max}} \arg\min_{\tau_p\in R}\left\|G_{a,\sigma,\tau_p} - S_p\right\|_2^2\right), \quad (2)$$

where $p_{\max}$ is the total number of measurement points and $v_{\max}$ is the total number of views.

The resulting Gaussian function for the X3-D display with $a = 0.51$, $\sigma = 7.49$ is shown in Fig. 9(a), along with $S_{12}(v)$. By sampling this curve for integer values of v and $\tau_p = v$, one can get the visibility of each view for observation point p. By replacing the view numbers in the interleaving map with their visibilities, one can obtain the visibility of each subpixel for observation point p.

Since the visibility curve is the same for all observation positions, and the optimal observation points are equi-

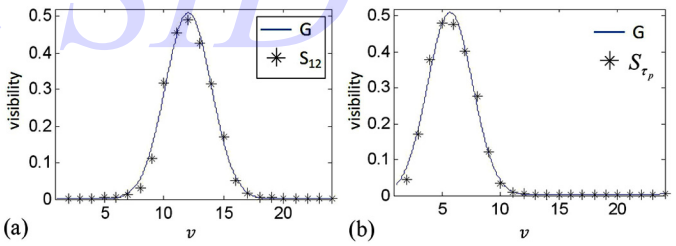


FIGURE 9 — Modeling angular visibility: (a) shape of derived Gaussian curve (G) with measurements for observation point 12 (S_{12}) and (b) predicted visibility of all elements (G) between observation points 5 and 6, with measurements done in that point (S_{τ_p}).

spaced, we can assume that the visibility of one view from different observation points follows the same function as visibility or different views from one observation point. In that case, one can estimate the visibility between two optimal observation positions by choosing fractional τ_p. As an example, $G(v)$ for $\tau_p = 5.5$ is shown in Fig. 9 along with actual measurements performed from an observation point that lies between points 5 and 6.

5 Performance analysis of a multi-view display in the frequency domain

Based on the measurements described in the previous sections, we can determine the visibility of every pixel on the LCD matrix from an arbitrary observation point. At the same time, due to the structure of the display, only a fraction of the pixels will be visible from one observation point (view). In addition, there are gaps between the visible subpixels in one view. Thus, when visualizing full-size images (*i.e.*, images with resolution equal to the LCD matrix resolution), there are two effects involved: aliasing due to the picking up of subpixels on nonrectangular grid and imaging, due to the presence of gaps. Aliasing can be fully tackled by an anti-aliasing prefilter. Imaging is usually tackled by an anti-imaging post-filter. Because the imaging is created by the physical structure of the display, it is impossible to impose a post-filter. However, the effect can be partially mitigated by a pre-filter. In order to determine the properties of the required 2-D filter, and consequently have the best possible representation of images on the display (minimizing aliasing, imaging, and ghosting), it is necessary to determine the performance of the display in the frequency domain; that is, we have to know which frequency components in the image we can keep (ones that will be properly represented on the screen), and which ones we have to attenuate (remove) as potential causes of distortions.

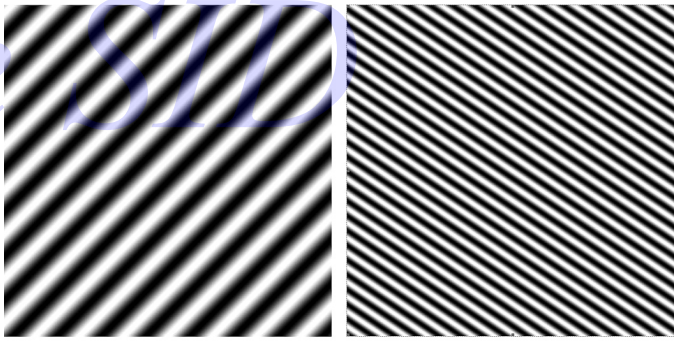
Determining the performance of the display in the frequency domain is challenging. Based on the interleaving pattern determined in Sec. 3 and the display model given in Fig. 1(b), one could derive an analytical expression of the frequency-domain behavior of the display. However, this theoretical approach does not take into account several effects occurring on the display. First, the visible subpixels are not on a rectangular grid (see Sec. 3). Second, every visible subpixel is seen with a different intensity (see Sec. 4). Finally, due to the masking effects, the pixels have a nonrectangular shape.[16–18] The combination of these effects creates an intensity distribution map, which alters the brightness of each image element in a non-linear fashion. Assuming that we could somehow model all non-linear effects, the analytical approach would become mathematically very demanding. Moreover, due to the fact that in the general case we do not know the exact properties of all parts of the screen (optical layer, slant of the barrier, thickness, *etc.*) in the theoretical approach many tradeoffs would have to be made, thereby ending with a frequency response that might or might not describe the display well enough. In order to achieve a practical solution, it turns out that for deriving the frequency response of the display, it is much more convenient to use a measurement-based approach, as described in this section. In the derived frequency response, we will denote as *passband* the region in the frequency domain containing frequencies that are properly represented on the screen. All other regions will be denoted as *stopband*.

The main idea in the proposed approach is to generate several images containing signals with various known frequencies, visualize them on the display, and then compare the output of the display with the input images. We illustrate the procedure for the X3-D display, but the approach itself is perfectly applicable for any other multi-view display.

5.1.1 Preparing test images

The first step in measuring the frequency response of the display is to generate appropriate test images. For this purpose, we prepared several hundred images, each of them being a pattern of a fixed known frequency. Two of these images for frequencies $(f_x, f_y) = (0.1, 0.1)$ and $(f_x, f_y) = (0.2, -0.3)$ are shown in Fig. 10 with the corresponding spectra shown as contour plots in Fig. 11. Here, f_x and f_y refer to frequencies along the x and y axis, respectively. The frequencies are normalized to $f_s/2 = 1$, with f_s being the sampling frequency. We can see in Fig. 11 that each of these signals has distinct peaks in the spectra [the peak at $(f_x, f_y) = (0, 0)$ is the DC component and should be ignored]. The motivation for using such images lies in the fact that by knowing exactly what image we sent to the display, based on what we see on the display, we can determine the properties of the display in the frequency domain.

We assume no *a priori* knowledge about the display properties. Therefore, the test images must be generated for all sets of frequencies (f_x, f_y) with $f_x \in [0, 1]$ and $f_y \in [-1, 1]$. The signals for $f_x \in [-1, 0)$ and $f_y \in [-1, 1]$ can be easily reconstructed by taking into account the symmetry properties of the spectra of real signals. In order to obtain a precise frequency response of the display, we have to use a very dense grid of frequencies. However, a very dense grid



(a) (b)

FIGURE 10 — Example of input (test) images: (a) $(f_x, f_y) = (0.1, 0.1)$ and (b) $(f_x, f_y) = (0.2, -0.3)$.

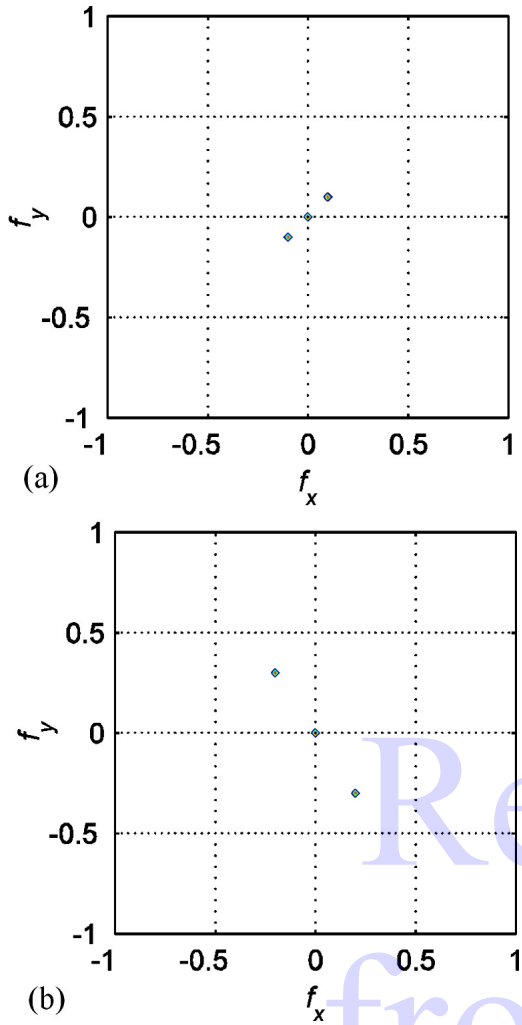


FIGURE 11 — Example of input images – Spectra (contour): (a) (f_x, f_y) = (0.1, 0.1) and (b) (f_x, f_y) = (0.2, –0.3).

means a high number of images which, in turn, will require a lot of measurements. Therefore, we suggest that first a larger grid is used, *e.g.*, $\Delta_f \geq 0.1$, to roughly determine the properties of the screen and then repeat the measurement with a denser grid, *e.g.*, $\Delta_f \approx 0.01$, in the regions around the edges of the passband. In this paper, for the X3-D display, we used the step $\Delta_f = 0.025$.

5.1.2 Measurements

The second step in measuring the frequency response of the display is to visualize the above described input images on the display and take photos of the screen by using a high-resolution digital camera. The photos were taken from a distance of ≈40 cm from the screen. Although the distance is not critical, the camera should not be put too close to the screen in order to avoid interference between multiple views. The photos taken by the camera will be hereafter referred to as output images. As an example, for the input images shown in Fig. 10, the output images are shown in Fig. 12 (the images have been enhanced for clarity) with the

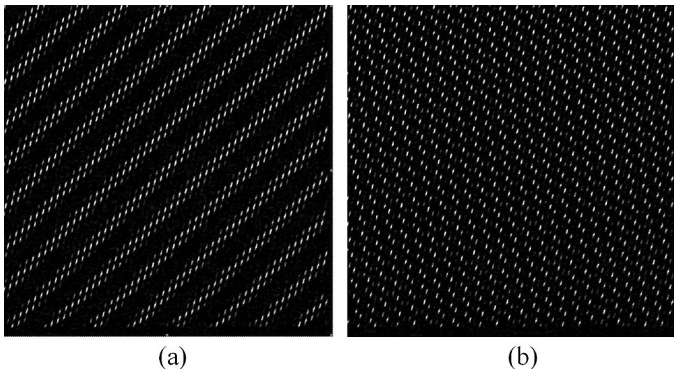


FIGURE 12 — Example of output images (photos taken from the display): (a) (f_x, f_y) = (0.1, 0.1) and (b) (f_x, f_y) = (0.2, –0.3).

corresponding spectra given as contour plots in Fig. 13. Please note that due to the fact that k, the ratio between pixels on the display and pixels in the output images, is less than one, after evaluating the spectra, the frequencies have to be rescaled by the factor $1/k$. This scaling has been already included in Fig. 13.

Several observations can be made based on these measurements. First, although each of the input images contains only a single frequency component, the output images contain many distinct frequency components. This is mainly due to the aliasing and imaging effects of the display. This is modeled in Fig. 13 through down-sampling and up-sampling blocks.

Second, although there are many high frequency distortions in the output image due to imaging, which is, in turn, due to the physical gaps between visible subpixels, those imaging components are partially suppressed by the human visual system. Therefore, we are still able to see properly the input signal on the display as long as the input signal contains only sufficiently low-frequency components. This is illustrated in Fig. 12(a) and Fig. 12(b). In the first figure, the diagonal lines are still seen even if they are heavily broken, but in the second figure, beside the barely visible diagonal lines from Fig. 10(b), many other lines are also seen and therefore we cannot properly identify the input signal.

Third, the display introduces non-linear distortions, as illustrated in Fig. 14. This figure shows the spectra along the

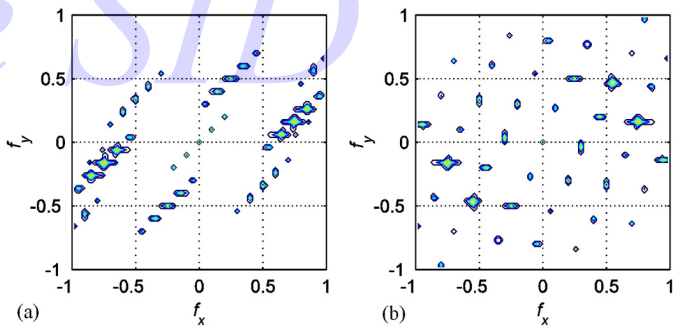


FIGURE 13 — Example of output images – Spectra (contour): (a) (f_x, f_y) = (0.1, 0.1) and (b) (f_x, f_y) = (0.2, –0.3).

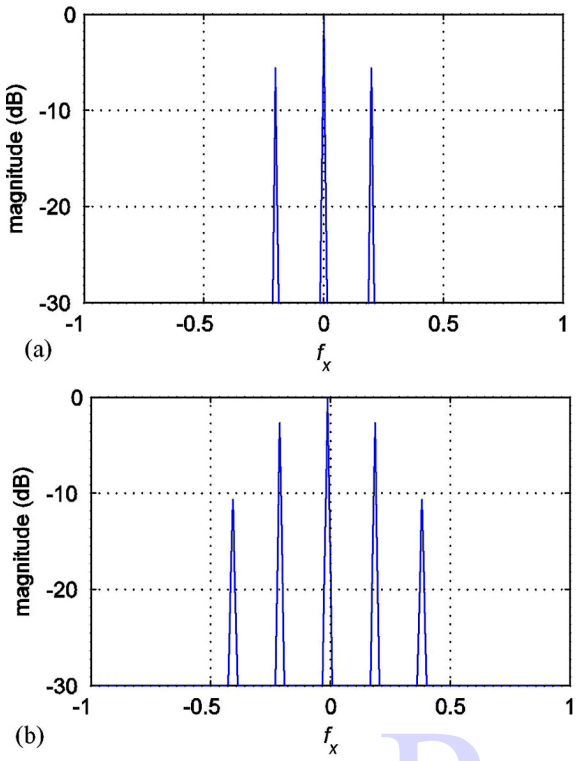


FIGURE 14 — Non-linear distortions – Spectra along *x*-axis for signal $(f_x, f_y) = (0.2, 0)$: (a) Input image and (b) output image.

x axis for the input signal $(f_x, f_y) = (0.2, 0)$ and the corresponding output signal. Although the input signal has only one spectral component at $f_x = \pm 0.2$, the output signal also contains harmonics at $f_x = \pm 0.4$ that are approximately 6–8 dB lower than the main spectral component.

5.1.3 Evaluating the frequency response of the display

The third and final step in measuring the frequency response of the display consists of comparing the spectra of the above derived input and output images. In order to suppress the noise in the output image, we threshold the spectrum of each image to –30 dB below the strongest frequency component. This threshold was experimentally chosen for the X3-D display, but we believe that it can be used for any other display having an 8-bit image depth.

The criteria for determining if a given frequency component passes through the system properly or if it is distorted due to the aliasing and imaging errors was the following: for every input signal of frequency (f_{x0}, f_{y0}), we checked if the contributing aliasing/imaging components contain frequency components that are inside a circle with radius

$$r_0 = \sqrt{f_{x_0}^2 + f_{y_0}^2}, \tag{3}$$

that is, if there are signals with a lower frequency than the one used at the input. In all cases we ignored the DC component. The motivation behind this criterion is two-fold.

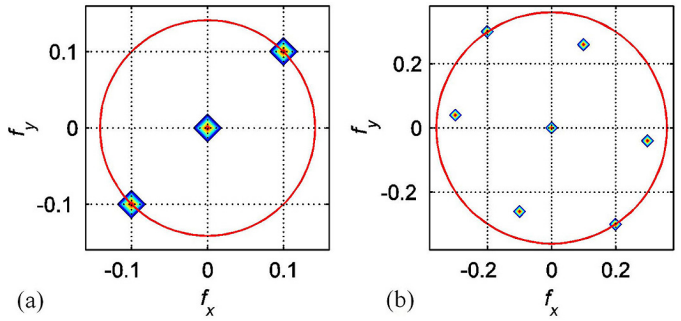


FIGURE 15 — Spectra of output images: (a) $(f_{x0}, f_{y0}) = (0.1, 0.1)$, $r_0 = 0.14$ and (b) $(f_{x0}, f_{y0}) = (0.2, -0.3)$, $r_0 = 0.36$.

First, according to the sampling theory, aliasing will occur once the input signal is greater than half of the sampling frequency. The frequency of the aliased component is lower than the frequency of the signal itself. Therefore, the above criteria effectively checks whether aliasing occurred or not. The second motivation lies in the fact that low-frequency errors (like the ones caused by aliasing) are much more visually annoying when occurring on the screen than high-frequency ones (*e.g.*, imaging). Consequently, we define the passband of the display as the region of all test frequencies which cause no additional frequency components inside radius r_0 and define the stopband everywhere else.

As an example, the magnified versions of Figs. 13(a) and 13(b) are shown in Figs. 15(a) and 15(b), respectively. In these figures, only the part containing frequencies smaller than r_0 (represented by the circle) is shown. As seen from the figures, in the first example there are no spectral components that are of a lower frequency than the one used for generating the input image, and therefore the signal of this frequency would be in the passband. In the second example, the output image considerably differs from the one sent to the display due to the aliasing errors. In the spectral domain this can be noticed by the presence of several frequency components that are inside the circle with radius r_0. There is no point in trying to represent images containing such input frequencies on the display under consideration.

By applying the above criteria to all the used input/output images, the passband and stopband of the display are classified as given in Fig. 16(a). In this figure, the passband is represented by dots. Due to measurement errors, the passband region is not continuous. It can be easily smoothened by applying a 5×5 median filter. The final frequency response for the X3-D display is shown in Fig. 16(b). As expected, the frequency response of the display given in Fig. 16(b) shows that the display is able to represent signals containing low frequencies. By following our methodology, one can obtain the passband region for an arbitrary display. This response can be used for deriving anti-aliasing filters to be applied to images before visualizing them on the display. Applying such filters will remove moiré artifacts and make masking artifacts less visible. An approach for designing efficient anti-aliasing filters for multi-view displays is discussed in Ref. 14.

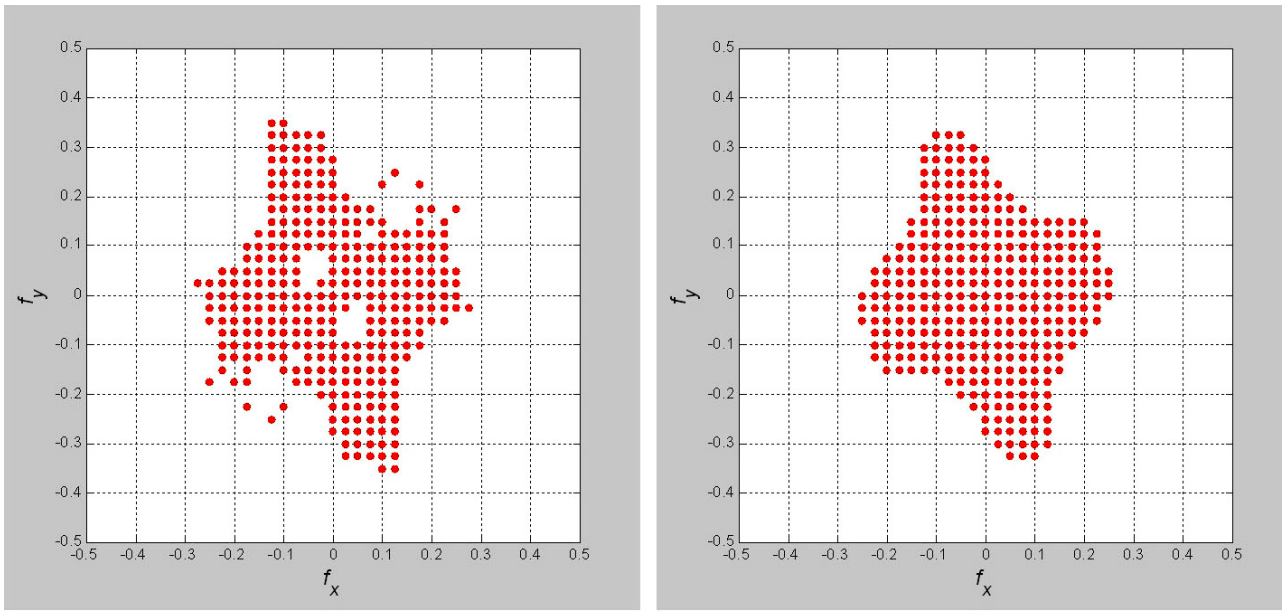


FIGURE 16 — Frequency response of the display: (a) passband region estimation based on measurements and (b) passband region after filtering.

6 Conclusions

In order to understand the reasons for artifacts in multi-view displays, we modeled the effect of optical layer on the underlying TFT-LCD image. Based on this model, we explained the reasons for some common artifacts. We identified which visual properties of a multi-view display are needed to design image-processing algorithms for artifact mitigation – namely, interleaving pattern, angular visibility of subpixels, and display performance in frequency domain.

We described a methodology for measuring and modeling these parameters, which does not require precisely positioned laboratory equipment. The methodology is simple, yet effective, and allows the end user to derive the number of images needed, the algorithm for interleaving them, and the pre-filter which would optimize the visual quality of these images for a given multi-view display.

To exemplify our methodology, we presented measurement results for one multi-view display. The precision of the model that we derived from the measurements is sufficient for it to be used in visual optimization algorithms.[14,19]

Acknowledgment

This work was supported by the EC within FP7 (Grant 216503 with acronym MOBILE3DTV) and by the Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006–2011).

References

1 S. Pastoor, "3-D displays," in: *3-D Video Communication*, Schreer, Kauff, and Sikora, eds. (Wiley & Sons, 2005).

2 J.-Y. Son and B. Javidi, "Three-dimensional imaging methods based on multiview images," *J. Display Technol.* **1**, 125– (2005).

3 J.-Y. Son *et al.*, "Recent developments in 3-D imaging Technologies," *J. Display Technol.* **99**, 1–10 (2010).

4 N. Dodgson, "Autostereoscopic 3-D Displays," *Computer* **38**, No. 8, 31–36 (Aug. 2005).

5 M. Salmimaa and T. Jarvenpaa, "Optical characterization of autostereoscopic 3-D displays," *J. Soc. Info. Display* **16**, 825 (2008).

6 C. Moller and A. Travis, "Correcting interperspective aliasing in autostereoscopic displays," *IEEE Trans. Visualization and Computer Graphics* **11**, No. 2, 228–236 (March 2005).

7 A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4-D-vision," *Proc. SPIE Photonics West 2002: Electronic Imaging* **4660**, 212–221 (2002).

8 V. Berkel and J. Clarke, "Characterisation and optimisation of 3-D-LCD module design," *Proc. SPIE* **2653**, *Stereoscopic Displays and Virtual Reality Systems IV* (Fisher, Merritt, Bolas, eds.), 179–186 (May 1997).

9 J. Konrad and P. Agniel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Process* **15**, 128–140 (Jan. 2006).

10 M. Salmimaa and T. Jarvenpaa, "3-D crosstalk and luminance uniformity from angular luminance profiles of multiview autostereoscopic 3-D displays," *J. Soc. Info. Display* **16**, 1033 (2008).

11 V. Saveljev *et al.*, "Moiré minimization condition in three-dimensional image displays," *J. Display Technol.* **1**, 347 (2005).

12 J. Hakkinen *et al.*, "Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width," *J. Soc. Info. Display* **17**, 433 (2009).

13 M. Zwicker *et al.*, "Antialiasing for automultiscopic 3-D displays," *ACM SIGGRAPH 2006 Sketches*, 107 (2006).

14 A. Boev *et al.*, "Anti-aliasing filtering of 2-D images for multi-view autostereoscopic displays," *Proc. 2009 International Workshop on Local and Non-Local Approximation in Image Processing*, LNLA (2009).

15 M. Becker, "Display reflectance: Basics, measurement, and rating," *J. Soc. Info. Display* **14**, 1003 (2006).

16 G. Woodgate and J. Harrold, "Efficiency analysis for multi-view spatially multiplexed autostereoscopic 2-D/3-D displays," *J. Soc. Info. Display* **15**, 873 (2007).

17 W. Mphepö *et al.*, "Enhancing the brightness of parallax barrier based 3-D flat panel mobile displays without compromising power consumption," *J. Display Technol.* **6**, No. 2, 60–64 (2010).

18 M. Krijn *et al.*, "2-D/3-D displays based on switchable lenticulars," *J. Soc. Info. Display* **16**, 847 (2008).

19 A. Boev *et al.*, "GPU-based algorithms for optimized visualization and crosstalk mitigation on a multiview display," *Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX* **6803** (2008).

20 P. Boher *et al.*, "A new way to characterize autostereoscopic 3-D displays using Fourier optics instrument," in: *Proc. SPIE, Stereoscopic displays and applications XX* (2008).
21 R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn., ISBN: 0521540518 (Cambridge University Press, March 2006).
22 P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *Proc. ACM SIGGRAPH* (1997).

Atanas Boev received his M.S. degree in 2001 from Technical University of Varna. From 2001 to 2002, we worked as a researcher in the Institute of Communication Engineering in Tampere University of Technology. In 2003 and 2004, he was Marie Curie Research Fellow in the Department of Signal Processing at the same university. From 2005 to the present day, he works as a researcher in the same department. His research interests are subjective quality of stereoscopic video and and algorithms for optimized visualization on autostereoscopic displays.

Robert Bregović received his Diploma Engineer and his M.S. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, in 1994 and 1998, respectively, and his Doctor of Technology degree (with honors) in information technology from the Tampere University of Technology in 2003. From 1994 to 1998, he was an Assistant at the Department of Electronic Systems and Information Processing of the Faculty of Electrical Engineering and Computing, University of Zagreb. In 1999, he was a Visiting Researcher at the Tampere International Center for Signal Processing, Institute of Signal Processing Laboratory, Department of Information Technology, Tampere University of Technology. From January 2000 to August 2003, he was a Researcher, and since September 2003, he has been a Post-Doctoral Researcher at same laboratory. His research interests are in digital signal processing, especially in multirate signal processing, digital filterbanks, and in generating optimization procedures for designing digital filters and filterbanks for various applications as well as in optimizing DSP algorithms for digital-signal-processor implementations.

Atanas Gotchev received his M.Sc. degrees in communications engineering and in applied mathematics from Technical University of Sofia, Sofia, Bulgaria, in 1990 and 1992, respectively, his Ph.D. degree in communications engineering from Bulgarian Academy of Sciences, Sofia, Bulgaria, in 1996, and his Dr. Tech. degree from Tampere University of Technology, Tampere, Finland, in 2003. Currently, he is Senior Researcher in the Institute of Signal Processing, Tampere University of Technology. His research interests are in transform methods for signal, image, and video processing.

[P05] A. Boev, A. Gotchev, “Comparative study of autostereoscopic displays for mobile devices”, *Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V.* Edited by Akopian, David; Creutzburg, Reiner; Snoek, Cees G. M.; Sebe, Nicu; Kennedy, Lyndon. Proceedings of the SPIE, Volume 7881, pp. 78810B-78810B-12 (2011)

Correction: Section 3d, paragraph 2, line 7, word 7. Correct: “convergence”.

This page left blank

Comparative study of autostereoscopic displays for mobile devices

Atanas Boev, Atanas Gotchev
Department of Signal Processing, Tampere University of Technology,
P. O. Box 553, FI-33101 Tampere, Finland

ABSTRACT

We perform comparative analysis of the visual quality of multiple 3D displays – seven portable ones, and a large 3D television set. We discuss two groups of parameters that influence the perceived quality of mobile 3D displays. The first group is related with the optical parameters of the displays, such as crosstalk or size of sweet spots. The second group includes content related parameters, such as objective and subjective comfort disparity range, suitable for a given display. We identify eight important parameters to be measured, and for each parameter we present the measurement methodology, and give comparative results for each display. Finally, we discuss the possibility of each display to visualize downscaled stereoscopic HD content with sufficient visual quality.

Keywords: portable 3D displays, parallax barrier, lenticular sheet, light guide, HDDP, crosstalk, accommodation-convergence rivalry, comfort disparity range, subjective quality, downscaled stereoscopic content.

1. INTRODUCTION

After the success of "Avatar 3D" movie in late 2009, twenty-eight stereoscopic movies were released in 2010, with thirty-six titles expected to come in 2011 [1]. The large amount of available 3D content encouraged the companies to produce 3D capable computers and television sets. In 2011, various models of mobile devices with 3D displays to are expected to become available. Examples of currently available devices with 3D displays, include a digital camera and photo frame [2] by Fujifilm, a mobile phone by Sharp [3]. Novel 3D-capable devices such as a game console by Nintendo [4], mobile phones by LG [5] and HTC [6], and a tablet by LG [5] are expected in 2011.

One unresolved question in the deployment of 3D-enabled mobile devices is whether the available 3D content will be suitable for the various mobile 3D displays, and to what extent some post-processing of the content will be needed. In order to select 3D display module, the vendors of mobile devices need to know how to compare the visual quality of such displays. In order to produce optimized 3D scenes, the content producers need to know what disparity range is suitable for a given display.

There are several studies on estimating the visual quality of 3D displays. Some studies propose analytical derivations based on knowledge of display properties [7, 8, 9], other studies measure the optical parameters of the displays [10, 11] or perform subjective tests [12, 13, 14]. Neither of this approaches is universally applicable, as display properties might not be known to the user, optical parameters might not be directly related to the perceived quality, and subjective tests are time consuming and expensive.

In this paper, we perform comparative analysis of the quality of seven portable 3D displays, and one large 3D television set. We try to identify the important parameters, which would influence the perceived quality of the display. For each parameter, we present the measurement methodology, and make compassion between the results for each display. The paper is organized as follows: in the next section we explain the operation principles of 3D displays. In Section 3 we discuss sources of visual discomfort specific for handheld 3D displays while in Section 4 we describe the display models included in our comparison. In Section 5, display related quality parameters are compared, such as crosstalk, sweet spot size, apparent size and resolution at the optimal viewing distance while Section 6 is about content related parameters, such as objective and subjective disparity comfort zone. Conclusions are given in Section 7.

2. MOBILE 3D DISPLAYS – PRINCIPLE OF OPERATION

Most mobile 3D displays are *autostereoscopic*. Autostereoscopic displays can create binocular illusion of depth without requiring the observer to wear special glasses. They work by beaming different image towards each eye. Most autostereoscopic displays use TFT-LCD matrix for image formation [15, 16]. Additional optical filter mounted on top of the

Multimedia on Mobile Devices 2011; and Multimedia Content Access: Algorithms and Systems V, edited by David Akopian, Reiner Creutzburg, Cees G. M. Snoek, Nicu Sebe, Lyndon Kennedy, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7881, 78810B · © 2011 SPIE-IS&T · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.879225

screen makes the visibility of each TFT element a function of the observation angle. There are three major types of optical filters. Displays with *lenticular sheet* have an array of microlenses which redirects the light, as shown in Figure 1a. This type of optical layer allows high brightness of the displayed image, but cannot be disabled electronically, thus these displays operate in 3D mode only. *Parallax barrier* works by partially blocking the light traveling in certain directions, as depicted in Figure 1b. This type of optical barrier allows less light trough, which results in a darker image, but allows the display to switch between "2D" and "3D" mode by turning the barrier on and off. Parallax barrier is the most commonly used optical filter in mobile 3D displays. The third approach – *lightguide plus retardation film* is exclusively used in displays produced by 3M. In this approach, the angular visibility of each TFT element can be altered by changing the position of the backlight [17]. Such displays have two sources of backlight, positioned on both sides of the display, as shown in Figure 1c. The images intended for left and right eye are visualized in a temporally-multiplexed manner where left and right backlights alternate over time and the content of the TFT elements is changed synchronously, thus allowing different image to be visible depending on the observation angle.

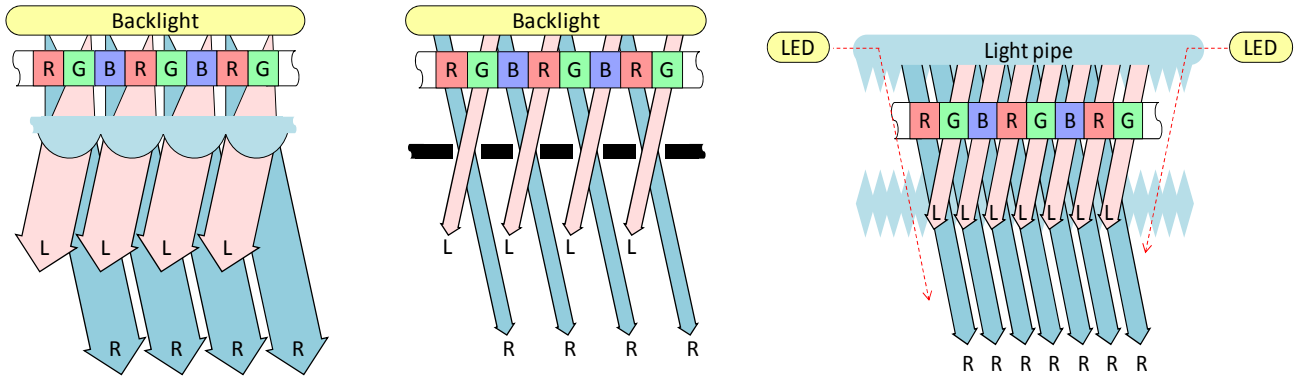


Figure 1. Different techniques for view separation in mobile 3D displays: a) leticular sheet, b) parallax barrier, c) lightguide plus retardation layer.

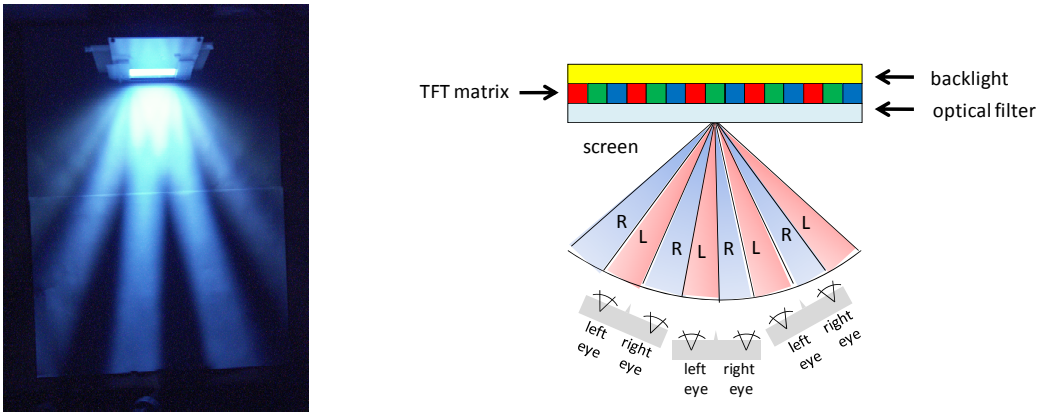


Figure 2. Views and their observation zones: a) observation zones of one view, b) observer positions for proper stereo effect.

The image visible on the autostereoscopic display from a given angle is called a *view* [15, 16]. As there is a tradeoff between the number of views and the resolution of one view, mobile autostereosopic display usually support two views only. These views contain the images intended for left and right eye of the observer, and thus are referred to as the *left view* and the *right view*. The difference in the horizontal position of an object in each view is called *disparity*, and it is responsible for the binocular illusion of depth. The illusory distance between the object and the display is referred to as *apparent depth*. Horizontal position of the object inside a view is measured as its distance between the object and the left edge of that view. Disparity is measured as the difference between the position of the object in the left and in the right view. Positive disparity creates the illusion of the object being behind the display, and negative disparity places the

object in front of the display. The range of positions from which a view is visible is called *visibility zone* of that view. A photo of the visibility zones of one view of a mobile 3D display is shown in Figure 2a. If a display has two views, their visibility zones alternate in horizontal direction. There are multiple positions (called *sweet spots*), from which an observer can perceive proper stereoscopic image as shown in Figure 2b. The procedure of mixing and mapping the images of both views to the TFT elements of the display is called *interleaving*, and the map, which determines if a TFT element belongs to the left or right view is known as *interleaving map*. We refer to interleaving maps, where all TFT elements on a row belong to the same view as *row-interleaved*, and to the ones, where columns of TFT elements belong to the same view - as *column-interleaved*.

3. SOURCES OF VISUAL DISCOMFORT IN MOBILE AUTOSTEREOSCOPIC DISPLAYS

a. Crosstalk

The observation zones of the two views are separated by a zone where neither of the views is predominantly visible. That region is sometimes referred to as *stereo-edge* and is shown in Figure 3a. At the sweet spot of a view, that view has maximum visibility while the visibility of the other view is suppressed as much as possible. Still, for any observation angle part of the light intended for one eye is also visible by the other. This process is usually modeled as *inter-channel crosstalk* and is expressed as the ratio between the luminance of one view to the luminance of the other [10], in percentages. The amount of crosstalk depends on the observation angle, and is lowest at the sweet spot of a view and is highest in the zone between two sweet spots. Subjective visual quality experiments described by Kooi et al [13] and Pastoor [14] suggest that inter-channel crosstalk of 25% is the maximum acceptable in stereoscopic image. Correspondingly, we define sweet spot as the area where crosstalk is less than 25%, as marked with gray in Figure 3a. The 3D displays are flat, and the observer sees different parts of the screen at a different angle as shown in Figure 3b. The positions, from which the same view is seen over the whole surface of the display have rhomboid shape, and are known as *viewing diamonds* [10, 11]. In our work we define the sweet spot of a view as the area within the viewing diamond where the crosstalk is less than 25%, as marked in gray in Figure 3b.

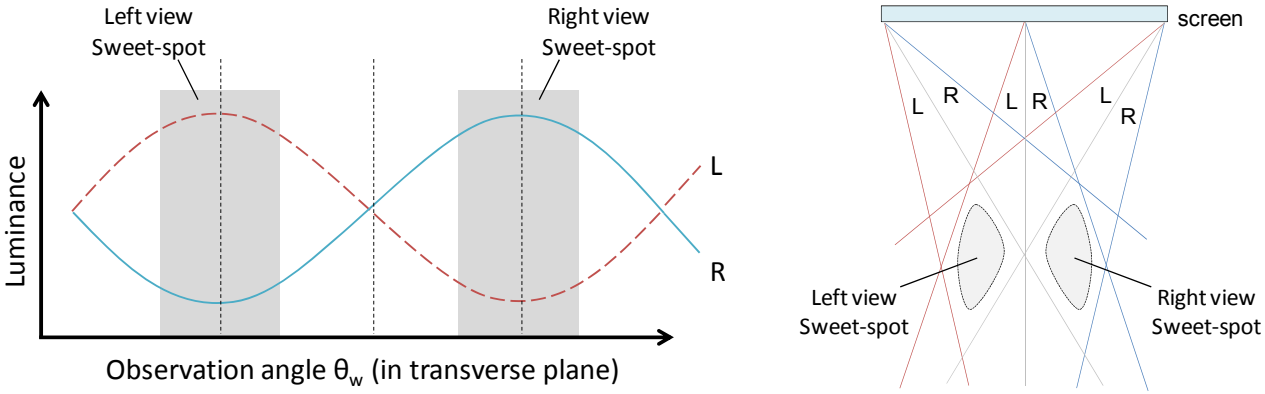
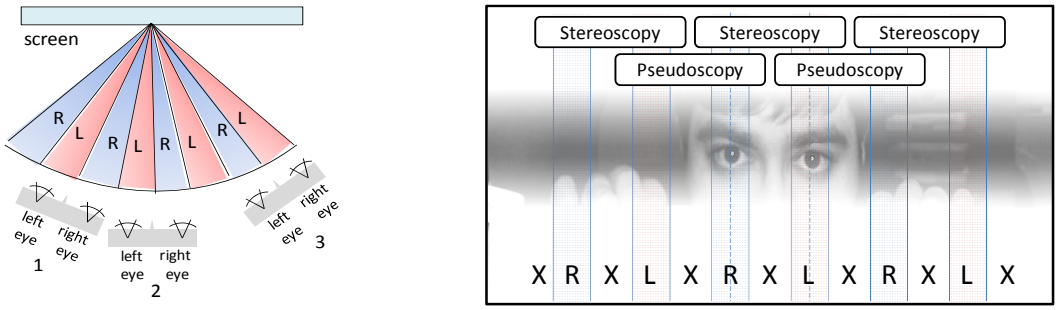


Figure 3. Crosstalk versus angle: a) visibility of each view and amount of crosstalk as function of the angle, b) position of the sweet spots.

b. Pseudoscopy

Pseudoscopy (*reverse stereo*) is the situation in which the eyes see the opposite views, i.e. the left eye sees the right view, and vice versa. Because of the repetitive order of the observation zones, there are multiple positions where the observer can perceive pseudoscopic image. For example, observers at positions marked with "1" and "2" in Figure 4a see proper stereo image, while the observer in position "3" experiences pseudoscopy. In pseudoscopic image, the binocular depth cues are reversed and the objects intended to appear in front of the display appear behind it and vice versa. In most cases this contradicts the depth suggested by other depth cues in the image (shadows, occlusion and parallax) and results in disturbing image [18]. Between the sweet spots and the areas producing pseudostopic image, there are zones with high crosstalk, where 3D effect is not visible – as marked with "X" in Figure 4b. We have found that

users of mobile 3D display intuitively change the observation angle in order to avoid zones with high crosstalk, however, avoiding pseudoscopic zones requires conscious effort.

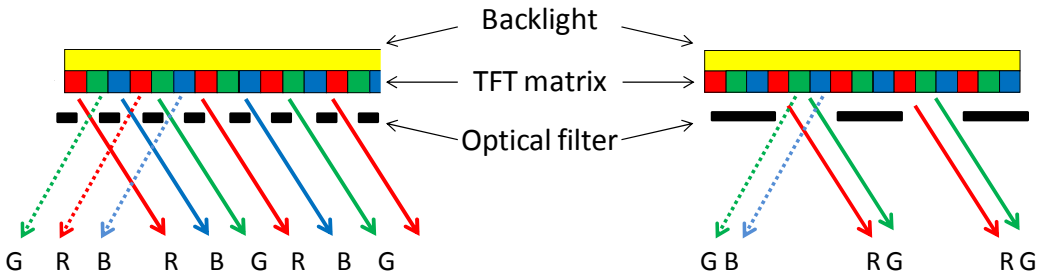


a) b)

Figure 4. Pseudoscopy in mobile 3D displays: a) proper (1 and 2) and wrong (3) positions of the observer, b) zones there proper perception is possible.

c. Aliasing and color bleeding

Autostereoscopic displays are column-interleaved (neighboring TFT elements in the same row belong to different views), thus their horizontal resolution is two times lower that their vertical one. As shown in Figure 5, only half of the TFT elements are visible from an observation position. This is equivalent to horizontal downsampling of the image by a factor of 2. Stereoscopic images are usually prepared for displays with square pixel aspect ratio and need to be suitably pre-filtered before visualized on a 3D display. Otherwise the downsampling performed by the optical filter might create *aliasing*, which manifests itself as Moiré artifacts. Lightguide-based 3D displays make an exception, as they do not suffer from spatial aliasing. However, they are susceptible to temporal aliasing artifacts.

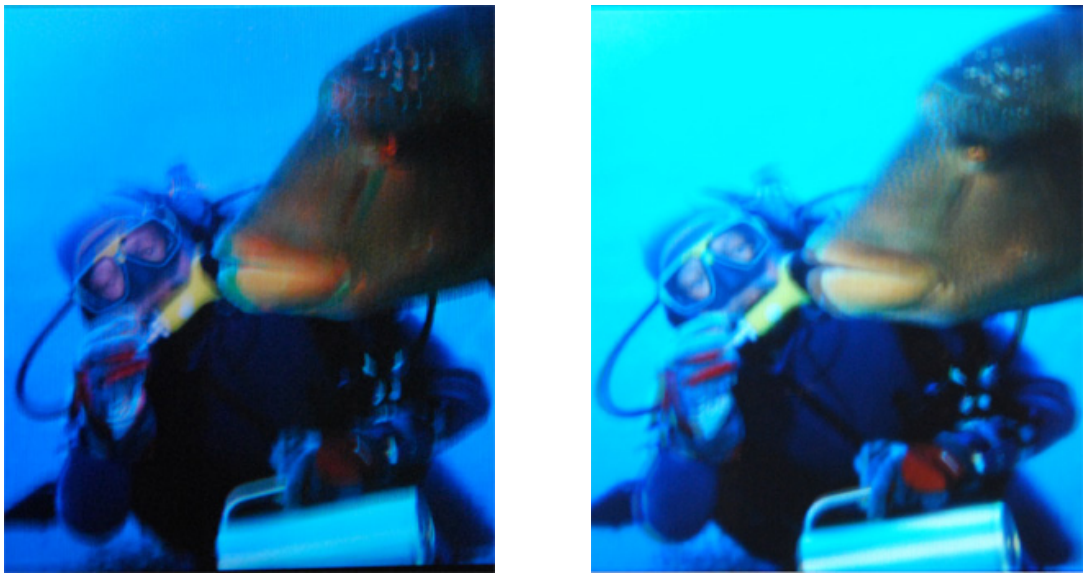


a) b)

Figure 5. Reasons for color bleeding: a) color balanced interleaving, b) interleaving, causing color misbalance.

The step size used to interleave TFT elements is a potential source of *color bleeding* artifacts. The elements of a color TFT display are of alternating red, green or blue color, and are also known as *sub-pixels*. If neighboring sub-pixels from the same row are of different color and the display is column interleaved, the color of the visible image becomes function of the observation angle. Some autostereoscopic displays are interleaved in a sub-pixel level (neighboring sub-pixels from the same row belong to a different view) as shown in Figure 5a. In such case for all observation angles the number of partially occluded sub-pixels is equally distributed along the three colors. In displays with other interleaving step size the amount of visible and partially occluded sub-pixels of certain color might prevail, which introduces color tint for some observation angles as exemplified in Figure 5b. These displays have reduced sweet spot width, since the optimal observation angle of each color channel is slightly different, and the zone which is optimal for all three colors is narrower. In [19] Uehara et al. proposed 3D display with horizontal double-density pixel (HDDP) arrangement, where the color of the sub-pixels change along columns, but the view assignment of sub-pixels change along rows. Displays with such topology do not exhibit color bleeding between the views. Additionally, because of the double pixel density in horizontal direction each view has square pixel aspect ratio which eliminates the most common reason for aliasing.

In Figure 6, one can see the difference between displays with different interleaving map, observed from an angle between two sweet spots. Both images have visible crosstalk, however, the image in Figure 6a also exhibits color bleeding, while the image in Figure 6b exhibits only crosstalk.



a) b)

Figure 6. Crosstalk versus color bleeding between views: a) crosstalk with balanced color distribution, b) crosstalk with additional color bleeding artifacts.

d. Factors limiting the disparity range

The perceived quality of a stereoscopic content is influenced by its maximum and minimum disparity. Content, optimized for large 3D display might not be suitable for smaller one or vice versa. There are two factors that influence the comfort of perception of content with given disparity range – *accommodation-convergence (A/C) rivalry*, and *divergent parallax*.

Human vision uses two oculomotor mechanisms to focus on a 3D object and adapt to its depth. One is *convergence*, in which both eyes perform inward or outward motion in order to bring the projection of the intended object to the foveae of both retinas. The other, called *accommodation* is the ability of each eye to change its focal power, so the projection of the object is focused on the retina. These two mechanisms are closely coupled, and the eyes automatically accommodate to the distance, suggested by the point of convergence [18]. In a natural 3D scene, such coupling increases the speed of accommodation and helps the convergence process by blurring the objects in front and behind the convergence point. However, on a stereoscopic display the conference and focal distances to an object differ. The distance to the converging point is influenced by object disparity, while the focal distance is always equal to the viewing distance, as shown in Figure 7a. This difference causes the objects with pronounced apparent depth to be perceived out-of-focus – an effect, known as *accommodation-convergence* (*A/C*) *rivalry*. In [20], A. Percival defines the combinations of focal and convergence distances, which allow clear vision. On a "focal distance" versus "convergence distance" plot, these combinations define so-called *zones of clear single vision* [21] (see Figure 7b). Beyond these zones, the A/C rivalry prevents eyes from converging, causing *diplopia* (double vision). However, inside the zones of clear single vision the observer still might experience A/C rivalry and see objects out of focus. In [22], the authors define so-called *Percival's zone of comfort*, which is approximately three times narrower than the zone of clear single vision, as shown in Figure 7b. Within that zone, A/C rivalry is negligible, which allows comfortable 3D perception [21, 22]. Notably, the smaller the focal distance is, the more pronounced A/C rivalry is, and smaller differences between focal and convergence distance could lead to uncomfortable 3D scene. As a consequence, the range of "comfortable" disparities is more limited for handheld 3D displays than for displays allowing greater viewing distance.

The inward and outward motion of the eyes is limited. Eyes can converge at distances ranging from about 5cm in front of the head to infinity. The eye muscles do not allow the eyes to look in divergent directions. The maximum disparity

that can be perceived is limited by the observer's inter-pupillary distance (IPD). If the disparity is larger, *divergent parallax* occurs, which is a disturbing, or potentially painful, experience [18]. This limitation is somewhat less pronounced in mobile 3D displays, as the mean IPD of 65mm corresponds to substantial part of the display width, the limits imposed by A/C rivalry occur for much lower values.

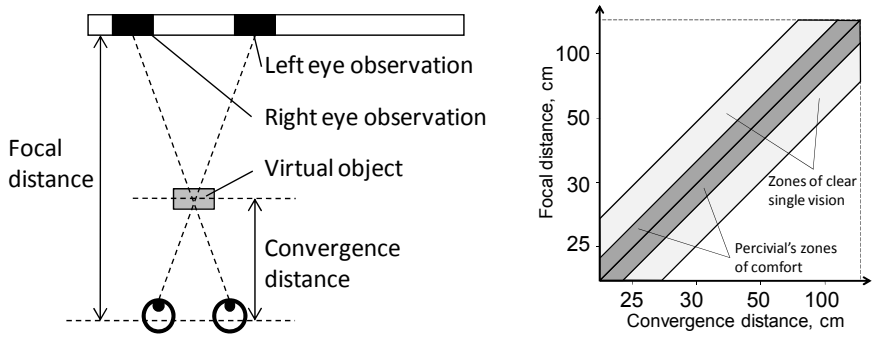


Figure 7. Accomodation-convergence rivalry: a) focal and apparent distance to of an object, b) zones of clear single vision and Persival's zones of comfort (adapted from [21]).

4. DISPLAY MODELS UNDER STUDY

In this work, we compare eight autostereosopic displays. The "HDDP" display is lentucular sheet display with HDDP pixel arrangement, produced by NEC LCD [19]. The "MI_L" and "MI_P" denote two orientations of a display produced by masterImage [23]. That display can operate in 3D-landscape and 3D-portrait mode by changing the direction of its parallax barrier. In either mode, the interleaving is per pixel. "MI_L" denotes the display operating in landscape mode, and "MI_P" – in portrait mode. "FC" is a portable stereoscopic camera produced by Fujifilm and equipped with 3D display [2]. The "FC" display uses light guide and retardation layer. Fujifilm also produces 3D photo frame which is labeled with "FF" in our comparison [2]. "SL" is a laptop with autostereoscopic 3D display produced by Sharp [24]. "V3D" is a prototype of a portable PC with 3D display, which is not in mass production. The "FF", "SL" and "V3D" displays are all 2D/3D switchable, column-interleaved on a sub-pixel level. For comparison, we have included two displays that work with polarized glasses. "AL" is a laptop with 3D display, produced by Acer, and "VUON" is large 3D television set with HDTV resolution.

Table 1 – display models under study.

Model	Description	Interleaving	Horizontal resolution (px)	Vertical resolution (px)	Width (cm)	Height (cm)
HDDP	3.2" display based on the lenticular HDDP technology by NEC	HDDP	427	240	6.9	3.9
MI_L	MB403M0117135 by Master Image (landscape 3D mode)	Column-interleaved, per pixel	800	480	9.3	5.6
MI_P	MB403M0117135 by Master Image (protrait 3D mode)	Column-interleaved, per pixel	480	800	5.6	9.3
FC	FinePix REAL 3D W1 camera by Fujifilm	Light guide	320	240	5.7	4.2
FF	FinePix REAL 3D V1 photo frame by Fujifilm	Column-interleaved, per sub-pixel	800	600	16.1	12.1
SL	Sharp AL3DU (with parallax barrier display)	Column-interleaved, per sub-pixel	1024	768	30.3	22.8
V3D	Portable computer with 3D display prototype	Column-interleaved, per sub-pixel	1024	600	10	5.8
AL	Acer AS5738DG-6165 laptop (polarized glasses)	Row interleaved	1366	768	34.3	19.3
VUON	Vuon E465SV 3D TV set by Hyundai (polarized glasses)	Row interleaved	1920	1080	91.5	57

5. DISPLAY RELATED PARAMETERS

a. Minimal crosstalk

The optical filter is rarely perfect, and even at the optimal observation position of a view there is still some residual crosstalk. For example, in the sweet spot of the left view (marked by "1" in Figure 8a and Figure 8b) the light of the

right view has some intensity I_{min}. At the same position, the left view is seen with maximum brightness, and has intensity I_{max}. One way to measure the crosstalk is to put vertical stripes in the left view and horizontal lines in the right view as proposed in [25]. The lines should consist of alternating black (minimum brightness) and white (maximum brightness) regions. In a sweet spot of a view either horizontal or vertical lines would be predominantly visible. From everywhere else the screen would appear with a square pattern as exemplified in Figure 8b. The appearance of the pattern can be used to identify the sweet spot of a view, and the crosstalk can be measured on a photograph of the display. After linearization of the camera response function of the camera making the photo (for example as described in [26]), the crosstalk can be measured by analyzing the brightness of four areas of the pattern as shown in Figure 8c. The darkest square in the pattern is where both views contain black pixels. Its intensity is denoted by I_{min}. The brightest part, where both views contain white pixels has intensity denoted by I_{max}. We refer to the view that is meant to be visible at the sweet spot where the measurement is done as the *current* view, and the other – as the *other view*. The intensity where the current view contains white pixels and the other view contains black pixels is denoted as I_c. The intensity of the part, where the current view has black pixels, and the other view is seen due to crosstalk is denoted as I_o. One should normalize I_C and I_O in a scale where I_{min}=0 and I_{max}=1, and then measure the crosstalk as the ratio between I_O and I_C in percentage. Altogether,

$$x_{3D} = \frac{I_O - I_{min}}{I_C - I_{min}} \cdot 100, \quad (1)$$

where x_{3D} is the crosstalk. According to [13], $x_{3D} < 5\%$ is considered unnoticeable, $10\% < x_{3D} < 25\%$ is noticeable, and crosstalk of more than 25% is rated as unacceptable. However, two of the displays in our comparison exhibited crosstalk higher than 25% - the FC display, which had crosstalk of 35%, and the AL display, which had crosstalk of 28% when used with general purpose polarized glasses. Subjectively, the "FC" display suffers from visible ghosting artifacts, but for natural content with highly textured areas 3D perception is still possible. Figure 9 gives a visual comparison between the amount of minimal crosstalk exhibited by our set of 3D displays and its perceptual impact.

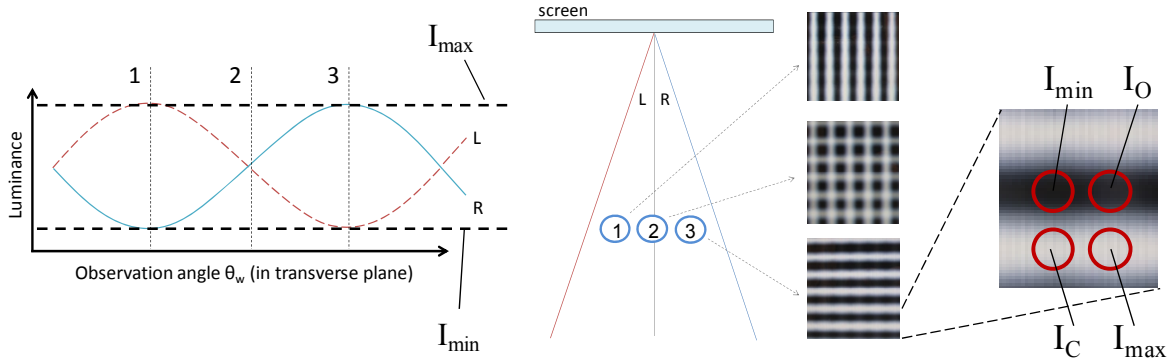


Figure 8. Minimal crosstalk: a) angular brightness, b) observation positions, c) visual example of minimal crosstalk.

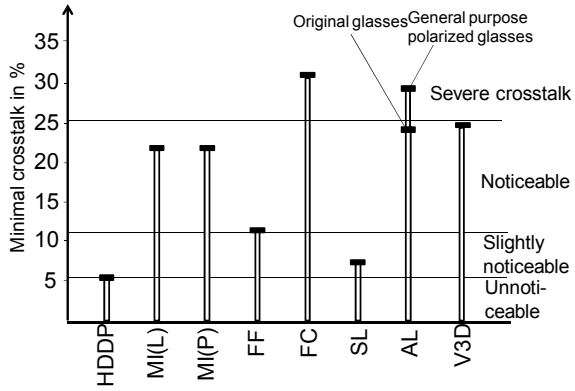


Figure 9. Minimal crosstalk per display model.

b. Optimal, minimal and maximal viewing distance

Most autostereoscopic displays can be used within a limited range of viewing distances. At one particular distance, TFT elements of the display are seen with a maximal brightness, and the visual separation between the views is optimal. This distance is called *optimal viewing distance* (OVD). The range of useful distances is limited by the need that both eyes appear inside the corresponding sweet spot. As shown in Figure 10a, the maximum (VD_{max}) and minimum (VD_{min}) usable distances depend on the IPD of the observer. In this work, we measure the sweet spots size for IPD=65mm. We define OVD as the distance, at which the crosstalk measured for each eye is minimal. We define VD_{min} to be the minimum distance, at which the crosstalk in each eye is lower than 25% and VD_{max} to be the maximum distance viewing distance which fulfils the same criterion. In Figure 10b, a graphical comparison between the optimal (long black bars), minimal and maximal (short black bars) viewing distances of various 3D displays is shown. The "AL" display uses polarized glasses, and its VD_{max} is beyond the usable observation distance. The VD_{min} of that display is limited by crosstalk visible along the top and bottom parts of the display. As OVD of "AL" display we used its nominal observation distance of 60cm.

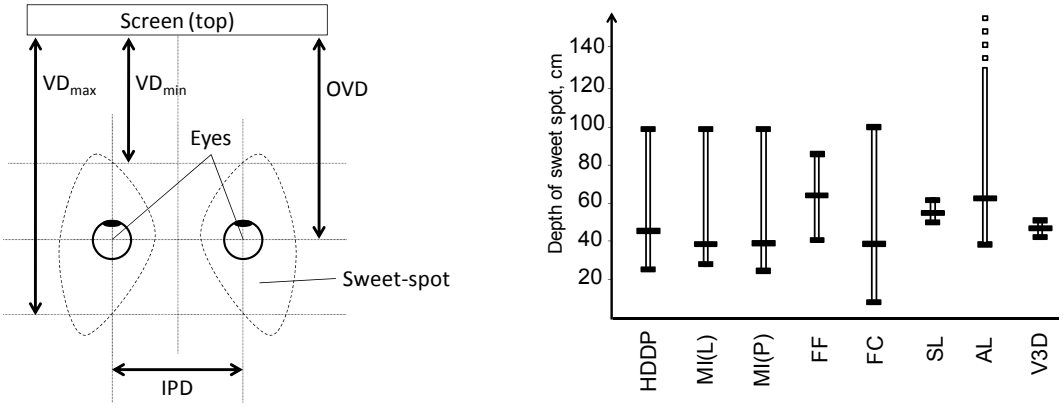
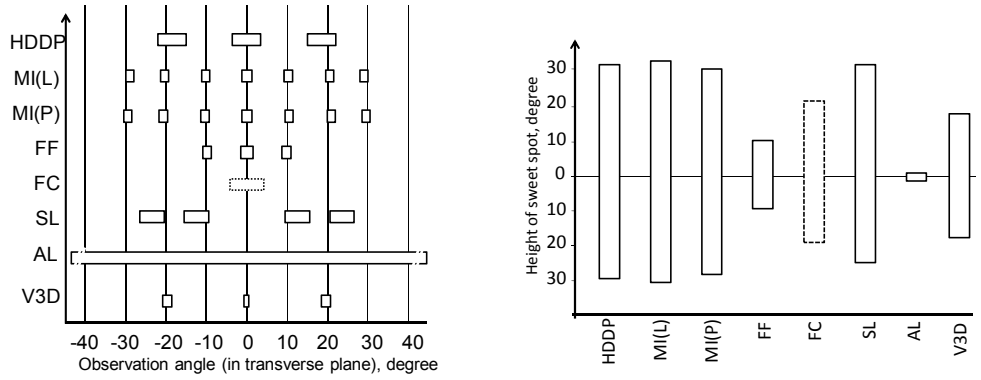


Figure 10. Optimal (long black bars), minimal and maximal (short black bars) viewing distances per display model.

Within the same setting, i.e. IPD=65mm and $x_{3D} < 25\%$, we measured the width and height of the sweet spots. The width of the sweet spots is measured in the transverse plane. Additionally, we excluded the pseudoscopic areas. The width of sweet spots for our set of displays is given in Figure 11a. The wider sweet spots are, the easier it would be for the observer to find the proper observation angle. Notably, the "AL" display works for all practical angles. Since the "FC" display has high minimal crosstalk, it does not fulfill the $x_{3D} < 25\%$ criterion for any observation angle. The width of the sweet spot for this display is measured for $x_{3D} < 35\%$ instead. The height of the sweet spots is measured in the saggital plane and is shown in Figure 11b. Most 3D displays have quite high sweet spots. Notably, the "AL" display, which has very wide sweet spot, is very sensitive to the vertical observation angle providing $x_{3D} < 25\%$ for vertical angles lower than 3 degrees up or down. Again, the "FF" display is measured for for $x_{3D} < 35\%$.



a) b)

Figure 11. a) width of sweetspot per display model, b) height of sweet spot per display model.

c. Angular size and resolution

When using a 2D display, the observer is free to choose an observation distance which would give the preferred trade-off between pixel density and field of view. This is not the case with 3D displays, which work best at their OVDs. In our comparison, we calculate the angle of view (AOV) of the display when observed at its OVD. We calculate the AOV as:

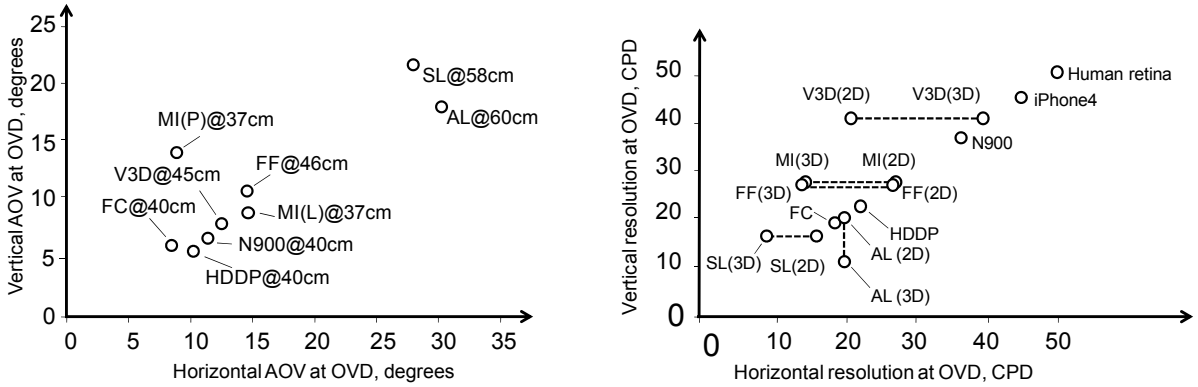
$$AOV_h = 2\arctan\frac{h}{2\cdot OVD}, \tag{2}$$

$$AOV_w = 2\arctan\frac{w}{2\cdot OVD}, \tag{3}$$

where AOV_h and AOV_w are the horizontal and vertical AOV, h and w are the horizontal and vertical size of the display. The calculations for our set of 3D displays are shown in Figure 12a. For comparison, we include the 2D displays used in Nokia N900 and Apple iPhone mobile devices. We measure pixel density in cycles per degree (CPD), generated by the display at its OVD. Since two pixels (black and white) are needed for one cycle, CPD is equal to the number of pixel pairs per centimeter that the display provides. After equivalent transformation, this is:

$$CPD = PPCM \cdot OVD \cdot \tan(0.5°), \tag{4}$$

where PPCM is the pixel density per centimeter for the display. The results are given in Figure 12b. For displays that can switch between 2D and 3D modes CPD is given separately for each case. Notably, the "HDDP" and "FC" displays have the same resolution in both modes. For comparison, the CPDs of N900 and iPhone are given, calculated for typical observation distance of 40cm. The resolution of the human retina for perfect 20/20 vision (50 CPD) is included as well.



a) b)

Figure 12. Angular size and resolution per display model measured at the optimal viewing distance: a) angular size, b) resolution in 2D and 3D mode.

6. CONTENT RELATED PARAMETERS

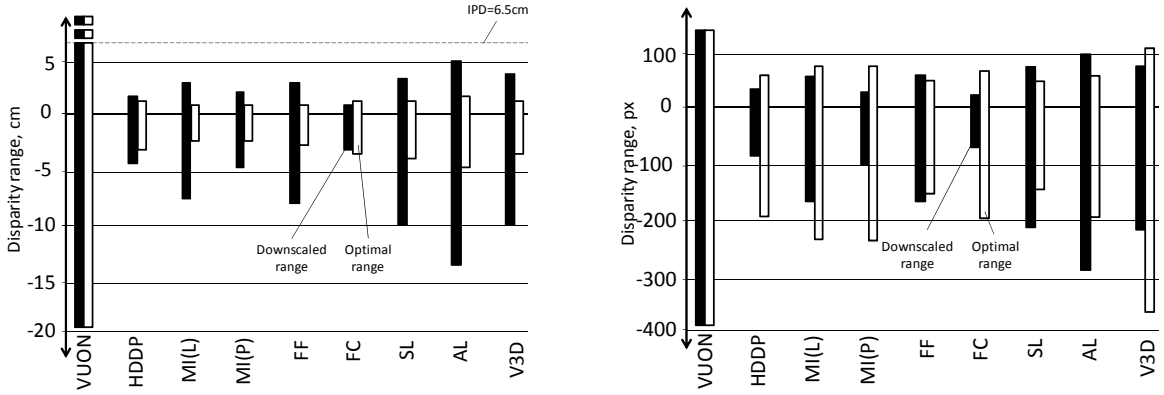
a. Objective comfort disparity range

There is a limit of disparity values that can be present in stereoscopic content in order for that content to be comfortably observed on a given stereoscopic display. We refer to that range to as *comfort disparity range*. The combined influence of A/C rivalry and divergent parallax determines the comfort disparity range of a given display. We calculated the disparity range, limited by these two factors for our set of 3D displays. In order to compare with disparity range of downscaled HDTV content, we calculated the ratio between the "VUON" display and each of the portable 3D displays. The ratios are listed in the second column of Table 2. We calculated the Persival's zone of comfort for the ODV of each display as explained in [21]. Then we converted the minimum and maximum apparent distance to disparity:

$$D_{min} = \frac{(OVD - l_{min})\cdot IPD}{OVD}; \tag{5}$$

$$D_{max} = \frac{(l_{max} - OVD)\cdot IPD}{OVD}; \tag{6}$$

where D_{min} and D_{max} are the minimum and maximum disparities in centimeters, l_{min} and l_{max} are the minimum and maximum distances to the convergence point, prescribed by the Percival's zone of comfort, also in centimeters. In our calculations, we used IPD=65mm. Using the optimal disparity range for the "VUON" display and the downscaling factors, we calculated the disparity range of a downscaled content for each display. Figure 13a gives the comparison between disparity range of downscaled content (black bars) and disparity range of display-optimized content (white bars), per display model, and disparity given in centimeters. In the figure, one can see that the comfort disparity range of all mobile 3D displays is insufficient to accommodate directly downscaled HDTV content. Scaling down stereoscopic image and observing it from closer distance increases the influence of A/C rivalry and content becomes uncomfortable to observe. However, mobile displays usually have higher pixel density than large TV sets, which results in further decreased disparity in centimeters for a given disparity in pixels. In Figure 13b we show comparison between downscaled range (black bars) versus optimal range (white bars) per display models, for disparities in pixels. For most mobile 3D displays, the additional downscaling caused by the higher pixel density compensates the decreased by A/C rivalry comfort disparity range, and allows the mobile display to accommodate the disparity range of a downscaled HDTV content.



a) b)

Figure 13. Downscaled (black) and optimal (white) disparity range per model: a) in cm, b) in pixels.

b. Subjective comfort disparity range

We have noted that the comfort disparity ranges predicted in Figure 13b do not coincide well with the subjective experience. Apparently, there are many additional factors that influence the comfort disparity range of a mobile 3D display – such as minimal crosstalk, optical quality, brightness and local contrast of the visualized content, to name a few. In order to determine the subjective comfort range of each display we performed small scale subjective test involving five observers. We prepared synthetic content with high contrast (white objects on black background) at different apparent depths. The participants had to choose the scene with the most pronounced, yet comfortably perceived depth range. We calculated the mean maximum and mean minimum value for each display. We used the subjective disparity range derived for the "VUON" display along with the downscaling factors, to calculate the disparity range of a downscaled 3D HDTV content. In Figure 14 we give a comparison between downscaled (black bars) and subjectively optimal (white bars) disparity ranges. In the most cases mobile 3D displays can accommodate downscaled stereoscopic HD content, with the exception of the "FC" display which suffers from high crosstalk. In some cases the optimal disparity range for a mobile display is sufficiently larger than the range of downscaled HD content. In Table 2 we show the relative difference between the two. All displays (except "FF") provide disparity range overhead, which allows the 3D HDTV content to be additionally repurposed, in order to increase the disparity range and provide extended range of apparent depth.

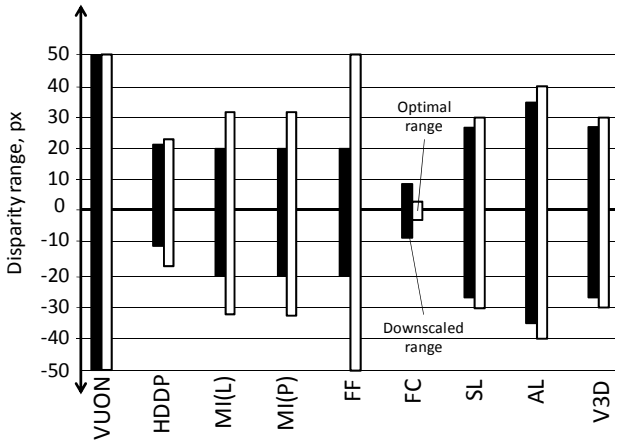


Figure 14, Downscaled (black) and optimal (white) disparity subjective comfort disparity range.

Table 2 – Rescaling factors and relative differences between optimized and downscaled content per display model.

Model	Rescaling factor (letterbox)	Negative disparity range	Positive disparity range
HDDP	0.23	54% more	98% more
MI_L	0.42	64% more	53.6% more
MI_P	0.25	64% more	53.6% more
FC	0.17	48% less	-52% less
FF	0.42	157% more	140% more
SL	0.53	20% more	12.5% more
AL	0.71	20% more	12.5% more
V3D	0.53	20% more	12.5% more

7. CONCLUSIONS

We discussed operation principles of mobile 3D displays. We listed the sources of visual discomfort and the most common artifacts exhibited by these displays. We performed comparative analysis of eight 3D displays, ranging from portable autostereoscopic to large stereoscopic television sets. We compared two groups of parameters – display related parameters, which describe the ability of a 3D display to faithfully reproduce a 3D scene, and content related parameters, which give information if a 3D scene with particular disparity range is suitable for a given stereoscopic display.

We have found that the properties of mobile 3D displays – such as resolution, optimal viewing distance, size of the sweet spot, comfort disparity range – vary a lot between display models, which makes it impossible to extrapolate "typical" properties of a mobile 3D display. According to our measurements, all portable 3D displays (except one) can accommodate the disparity range of downscaled stereoscopic content. In some cases, the optimal disparity range is sufficiently larger, which allows HD content to be additionally repurposed for a portable 3D display, in order to extend the apparent depth range of the content.

Finally, our experiments showed that calculations of the viewing geometry are not sufficient for precise prediction of the comfort disparity zone of a 3D display. In a future work we will study the influence of crosstalk and local contrast on the comfort disparity range of a given stereoscopic display.

ACKNOWLEDGEMENTS

This work was supported by the European Commission within FP7 (Grant 216503 with the acronym MOBILE3DTV).

REFERENCES

1 "List of 3D films", Wikipedia, available online at: http://en.wikipedia.org/wiki/List_of_3-D_films

2 "FinePix REAL 3D series", product brochure, Fujifilm, 2010, Available: http://fujifilm.co.uk/media/dContent/mediaCentre/Brochures/0_FinePix-Real-3D-catalogue.pdf

3 "Sharp Develops 3D Touchscreen LCD—Switchable Between 2D and 3D Modes—with Industry's Highest Brightness", Sharp press release, available online at: http://sharp-world.com/corporate/news/100402.html

4 "Nintendo 3DS Brings a Dimensional Shift to the World of Entertainment on March 27", Nintendo, Press release, available online at: http://www.nintendo.com/whatsnew/detail/Yc3LhjvhjnDIlUEBTdnVEg-cLSs0NPfJ

5 "LG Optimus Pad and LG Optimus 3D Phone Features Revealed", PlanetInsane news, available online at: http://www.planetinsane.com/lg-optimus-pad-and-lg-optimus-3d-phone-features-revealed/268788/

6 "HTC 3D phone", The Tech Journal, available online at: http://thetechjournal.com/electronics/mobile/htc-3d-phone.xhtml

7 Berkel, V. and Clarke, J., "Characterisation and optimisation of 3D-LCD module design", *Proc. SPIE 2653,* 179-186 (1997)

8 Konrad, J. and Agniel**,** P., "Subsampling models and anti-alias filters for 3-D automultiscopic displays," IEEE Trans. Image Process. 15, 128-140 (2006).

9 Saveljev, V., Son, J.-Y., Javidi, B., Kim, S.-K. and Kim, D., "Moiré Minimization Condition in Three-Dimensional Image Displays," J. Display Technol. 1, 347- 353 (2005).

10 Salmimaa, M. and Jarvenpaa, T., "Optical characterization of autostereoscopic 3-D displays", J. Soc. Inf. Display 16, 825-833 (2008).

11 Boher, P., Leroux, T., Bignon, T. and Collomb-Patton, V., "A new way to characterize auto-stereoscopic 3D displays using Fourier optics instrument", Proc. of SPIE 7237, 72370Z (2009).

12 Hakkinen, J., Takatalo, J., Kilpelainen, M., Salmimaa, M. and Nyman, G., "Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width," J. Soc. Inf. Display 17, 433-441 (2009).

13 Kooi, F. and Toet, A., "Visual comfort of binocular and 3D displays," Displays 25 (2-3), 99-108 (2004).

14 Pastoor, S., "Human factors of 3D images: Results of recent research at Heinrich-Hertz-Institut Berlin," Proceedings of IDW'95, 69-72 (1995).

15 Pastoor, S., "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

16 Dodgson, N., "Autostereoscopic 3D Displays," Computer 38(8), 31- 36 (2005).

17 3M Press release, "3M Enhances 3D Display Capabilities for Handheld Displays with 3D Optical Film," available online at: http://www.businesswire.com/news/3m/20091005005255/en/3M-Enhances-3D-Display-Capabilities-Handheld-Displays, 2009.

18 IJsselsteijn, W., Seuntiens, P. and Meesters, L., "Human factors of 3D displays", 3D Video Communication (Schreer, Kauff, Sikora, Edts.), Wiley 2005.

19 Uehara, S., Hiroya, T.., Kusanagi, H., Shigemura, K. and Asada, H., "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement",Proc. SPIE 6803, 68030O-68030O-8 (2008).

20 Percival, A. S., [The prescribing of spectacles], Bristol, J. Wright, (1920).

21 Hoffman, D., Girshick, A., Akeley, K. and Banks, M., "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue", Journal of Vision 8(3), 1–30 (2008).

22 Howard, I. P., & Rogers, B. J., [Seeing in depth], University of Toronto Press, Toronto, (2002).

23 MasterImage , 3D-LCD product brochure, available online at http://masterimage.co.kr/new_eng/data/masterimage.zip?pos=60

24 Sharp, Sharp Laboratories of Europe, http://www.sle.sharp.co.uk/research/optical_imaging/3d_research.php, 2010

25 Scala, V., "Cross-talk measurement for 3D dispalays," 3DTV-CON 2009*,* 4 pages (2009)

26 Debevec, P., Malik, J., "Recovering High Dynamic Range Radiance Maps from Photographs," Proc. ACM SIGGRAPH, (1997)

[P06] A. Boev, R. Bregovic, A. Gotchev, “Design of tuneable anti-aliasing filters for multiview displays”, *Stereoscopic Displays and Applications XXII, Proc. SPIE 7863*, 78630F (2011), DOI: 10.1117/12.873465

This page left blank

Design of tuneable anti-aliasing filters for multiview displays

Atanas Boev, Robert Bregovic, Atanas Gotchev
Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, 33101 Tampere, Finland

ABSTRACT

Multiview displays suffer from two common artifacts – Moiré, caused by aliasing, and ghosting artifacts caused by crosstalk. By measuring the angular brightness function of each TFT element we create so-called brightness mask, which allows us to simulate the display output for a given input image. We consider multiview display as image processing channel and model the artifacts as distortions of the input signal. We test the channel by using a set of signals with various frequency components as input, and analyzing the output in the frequency domain. We derive the so-called bandpass region of the display, where the distortions introduced to the input signals are under certain threshold. Then, we extend the simulations including input signals with varying disparity, and obtain multiple passbands – one for each disparity level. We approximate each passband with a rectangle and store the height and width of that rectangle in a table.

We propose an artifact mitigation framework which can be used for realtime processing of textures with known apparent depth. The framework gives the user ability to set so-called "3D-shapness" – a parameter, which controls the trade-off between visibility of details and presence of artifacts. The "3D-sharpness parameter determines what level of distortions is allowed in the final image, regardless of its disparity. The framework uses the approximated width and height of the passband areas in order to design an optimal (for the needed disparity and desired distortion level) anti-aliasing filter. We discuss a methodology for filter design, and show example results, based on measurements of an 8-view display.

Keywords: multiview displays, anti-aliasing filters, optical measurements, Moiré, ghosting artifacts, filter design, 3D-sharpness

1. INTRODUCTION

Multiview displays can create stereoscopic 3D effect without requiring the observer to wear specially designed glasses. They work by simultaneously visualizing a number of images, each one visible at different angle. Most often, a multiview display uses a TFT-LCD matrix for image creation, and an additional *optical layer* mounted on top, which redirects the light created by the TFT elements [1, 2]. The visibility of the TFT color components (also known as *subpixels*) becomes a function of the observation angle. From each angle, a group of subpixels is predominantly visible, and forms an image. Such image is called a *view*. If multiple observations of the same scene are properly assigned to the views of a multiview display, the stereoscopic 3D effect is created. The process of mapping multiple images to the subpixels of a multiview display is known as *interleaving*, and the map, which defines the correspondence between subpixels and view number is known as *interleaving map*. The interleaving map has the full resolution of the TFT-LCD matrix of the display, but most often it can be fully described by a smaller repetitive structure that we refer to as *interleaving pattern*.

The design of the optical filter involves a trade-off between number of views, resolution of a view and the quality of view reconstruction. Multiview displays are susceptible to a number of visual artifacts [3], but two are the most pronounced – Moiré and ghosting [3, 4, 5]. Mapping an image to the visible pixels of a view involves subsampling, which often happens on a non-rectangular grid. Failing to properly pre-filter the image creates aliasing, which manifests itself as Moiré artifacts. The effect is more complicated for objects with pronounced depth. In that case, observations of the same object appear in different horizontal position in each view. The horizontal offset between the observations is known as *disparity*. The disparity between the observations in different views creates the stereoscopic illusion of depth. In this work, we refer to the illusory distance to the object created by the stereoscopic effect as *apparent depth*. Positive disparity creates apparent depth behind the screen plane, and negative disparity creates apparent depth in front of the screen. In order to avoid banding and image flipping artifacts, the observation zones of the views are interspersed, and from each observation direction a number of views are simultaneously visible (albeit with different brightness). This effect is modeled as crosstalk between the views. The combination of crosstalk and disparity creates horizontally shifted, semi-visible replica of visualized object – an effect called *ghost images*, or ghosting.

Stereoscopic Displays and Applications XXII, edited by Andrew J. Woods, Nicolas S. Holliman, Neil A. Dodgson, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 7863, 78630F

Moiré artifacts are especially visible in images with high contrast and sharp details – such as text or GUI widgets. In 3D scenes, such content can have different apparent depth, for example depth of subtitles is rendered according to depth of the rest of the scene. The visibility of distortions varies with the frequency, orientation, and depth of the content. In previous work, we proposed a methodology for design of anti-aliasing filters based on the frequency performance of a multiview display [6]. These filters were optimal for 2D content with no apparent depth. Additionally, we have found that the filter that fully suppresses aliasing does not always give the best perceptual quality [7]. Some people prefer sharper-looking images at the expense of some Moiré artifacts. In this work, we extend our methodology towards design of anti-aliasing filters for content at different disparity levels. We discuss methodology for design of tuneable filters, which depend on two parameters – apparent depth and desired sharpness. The sharpness parameter is expressed in terms of signal-to-distortion ratio, which we claim to affect the visibility of aliasing in perceptually linear fashion, regardless of the apparent depth. Throughout the paper, we give measurement results for an 8-view multiview display as a practical example. The display we use for the examples is 23"-Multiview Display AD, produced by Opticality (formely X3D Corporation) to which we refer to as the *X3D-display* [8].

The paper is structured as follows. In the next section, we introduce the method for performance analysis in the frequency domain, which is used for obtaining the so-called passband region of the display. In Section 3 we study the passband region for input signals with varying apparent depth. In Section 4 we introduce the concept of perceived distortion, and show how to derive the passband region as a function of the distortion percentage. Then, in Section 5 we propose a combination of display measurements and image processing framework, that uses tunable filters for anti-aliasing of objects with given depth for desired sharpness for a given multiview display. Section 6 explains the design of the tunable filters used in this paper. Finally, in Section 7 we give examples for the performance of the filter, optimized for the X3D-display.

2. ESTIMATING THE DISPLAY PERFORMANCE IN THE FREQUENCY DOMAIN

2.1 Multiview display as image processing channel

Most often, the optical layer has a regular structure, and affects the underlying image in a non-uniform way. Therefore image details with certain density and orientation are more prone to distortion than others [12]. Notable exception are displays with random hole distribution [9] where all textures would be distorted equally despite of their frequency content. In order to study how various frequency components get deteriorated by Moiré and ghosting artifacts, we propose a model of a multiview display that considers the display as an image processing channel. The model follows the steps of rendering 2D texture with given apparent depth on a multiview display. Block diagram of the model is shown in Figure 1.

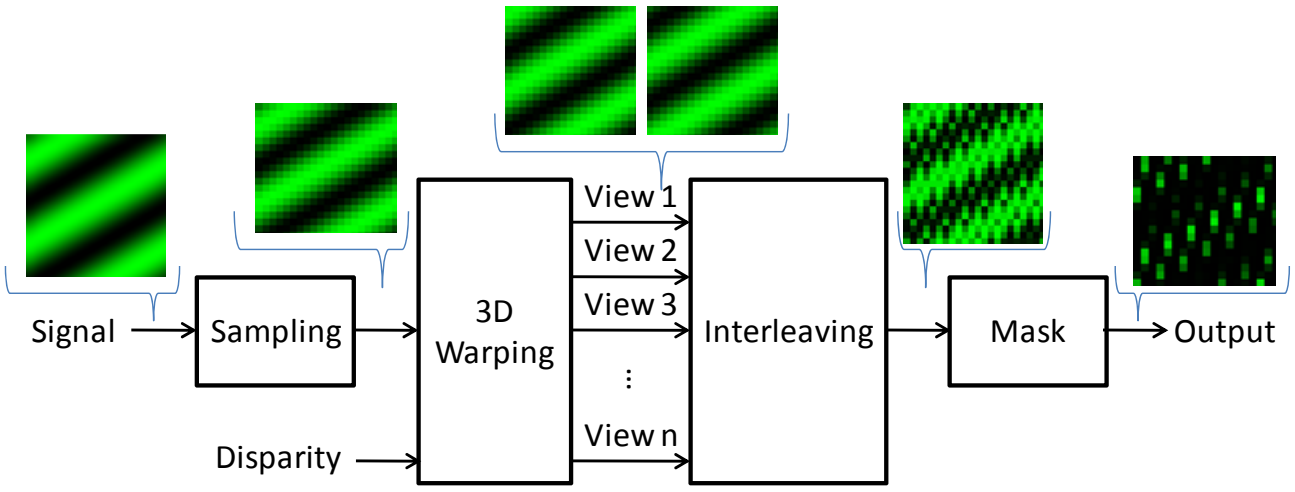


Figure 1, model of a multiview display as image processing channel. The model assumes that the input signal is 2D texture with known apparent depth.

The input to the model is a continuous 2D signal – e.g. text, GUI widget or texture patch. The signal is sampled with the resolution of the underlying TFT-LCD (in other words, with the resolution of the interleaving map). The next step is 3D

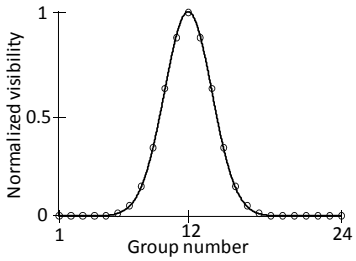
warping. In order to render the input signal at given apparent depth, shifted versions of the sampled signal are created and assigned to different views. After that, all views are interleaved into one compound image, as it is prescribed by the interleaving map of the multiview display. Since the compound image has the same resolution as the interleaving map, the interleaving process involves downsampling. From each source image only part of the pixels are included in the compound image. The compound image is shown on the TFT-LCD matrix and is transformed by the optical layer. The optical layer acts as a mask, which alters the brightness of each underlying TFT element as a function of the angle. In the ideal case, the visible image should be replica of the image in one of the views. In the real case, parts of the input image are missing, and parts of the images which belong to other views are partially visible.

2.2 Simulation of the display output

The interleaving pattern is provided to the end user, for example the pattern of the X3D-display is seen in Figure 2a. However, such patterns are often imprecise and deriving it experimentally might give more accurate results. In a previous work, we introduced a methodology for deriving the interleaving pattern of a multiview display, and for deriving the so-called angular visibility function of each TFT element of that display [10]. For X3D-display we found 24 distinctive groups of TFT elements with equivalent angular visibility, as opposed to 8 views as stated in the manual. Hence, one can use the X3D-display with 24 views, albeit with low resolution in each view and high crosstalk between the neighboring views.

	Column 1			Column 2			Column 3	
row	1R	1G	1B	2R	2G	2B	3R	3G
1	1	2	3	4	5	6	7	8
2	2	3	4	5	6	7	8	1
3	2	3	4	5	6	7	8	1
4	3	4	5	6	7	8	1	2
5	4	5	6	7	8	1	2	3
6	4	5	6	7	8	1	2	3
7	5	6	7	8	1	2	3	4
8	6	7	8	1	2	3	4	5
9	6	7	8	1	2	3	4	5
10	7	8	1	2	3	4	5	6
11	8	1	2	3	4	5	6	7
12	8	1	2	3	4	5	6	7

a)



b)

	Column 1			Column 2			Column 3	
row	1R	1G	1B	2R	2G	2B	3R	3G
1	0.1	0.13	0.45	0.74	0.55	0.14	0.1	0.1
2	0.11	0.25	0.73	0.66	0.19	0.11	0.1	0.09
3	0.16	0.65	0.71	0.34	0.12	0.1	0.1	0.1
4	0.45	0.74	0.55	0.14	0.1	0.1	0.1	0.13
5	0.73	0.66	0.19	0.11	0.1	0.09	0.11	0.25
6	0.71	0.34	0.12	0.1	0.1	0.1	0.16	0.65
7	0.55	0.14	0.1	0.1	0.1	0.13	0.45	0.74
8	0.19	0.11	0.1	0.09	0.11	0.25	0.73	0.66
9	0.12	0.1	0.1	0.1	0.16	0.65	0.71	0.34
10	0.1	0.1	0.1	0.13	0.45	0.74	0.55	0.14
11	0.1	0.09	0.11	0.25	0.73	0.66	0.19	0.11
12	0.1	0.1	0.16	0.65	0.71	0.34	0.12	0.1

c)

Figure 2, deriving the visibility mask of X3D-display: a) interleaving pattern, b) measured visibility per group of TFT elements, c) visibility pattern, that can be used as weighting mask for simulating the effect of the optical layer.

One of the commonly measured parameters of a multiview display is the angular brightness of a TFT element, which is the brightness of that element as a function of the observation angle. In this work we use so-called *angular visibility*, which is proportional to the angular brightness, but is normalized in the range between 0 and 1, where 1 is the maximum brightness of a given TFT element. For example, the angular visibility for the 24 groups of TFT elements as seen directly in front of X3D-display is given in Figure 2b. The angular visibility can be directly used as a weighting mask on the values of the interleaved image. This allows the appearance of a multiview display to be simulated for given interleaved image and observation angle. In our experiments, we simulate X3D-display as having 8 views, as given in the display manual, but use the precise weighting mask based on the derived 24 groups of TFT elements with similar angular visibility[10]. The weighting mask is shown in Figure 2c.

a) b) c)

Figure 3, simulation of the display output: a) input signal, b) simulated output, c) photograph input signal, visualized on X3D-display.

In our experiments, we simulate X3D display as it would be seen directly in front of the screen, at 150cm distance, which is the sweet-spot of view number 5. One of our test images can be seen in Figure 3a, and its corresponding simulated version can be seen in Figure 3b. For comparison, in Figure 3c we give photograph of the same image, visualized on the X3D-display.

2.3 Performance analysis

The most important question in connection with a multiview display is related to the proper representation of an image, that is, will an image be properly seen on the display or not? In order to answer this, we must determine which signal frequencies can be properly represented on the screen. In theory, we could use our knowledge about the interleaving pattern and the corresponding brightness (see Figure 2) to derive theoretical expressions that would describe the performance of the display in the frequency domain. Unfortunately, this is a mathematically very demanding task. In order to simplify the analysis, in this paper we will use the method introduced in [10] that was derived for analyzing the display based on a set of images obtained by photographing the display (measurements). From the method point of view, it does not matter if the processed image is a photo of the display or is simulated. Since, we can easily simulate the output of the display, the performance analysis in this paper becomes straightforward. It should be pointed out that the results obtained by simulation and the ones obtained by measurement can considerably differ from each other for a given display due to various effects in the display that are not modeled in the simulation. However, for presenting the basic concept of tunable filters introduced in this paper, simulation results will be sufficient. For completeness of this paper, next, we will briefly describe the method proposed in [10]. More detail about the method can be found there.

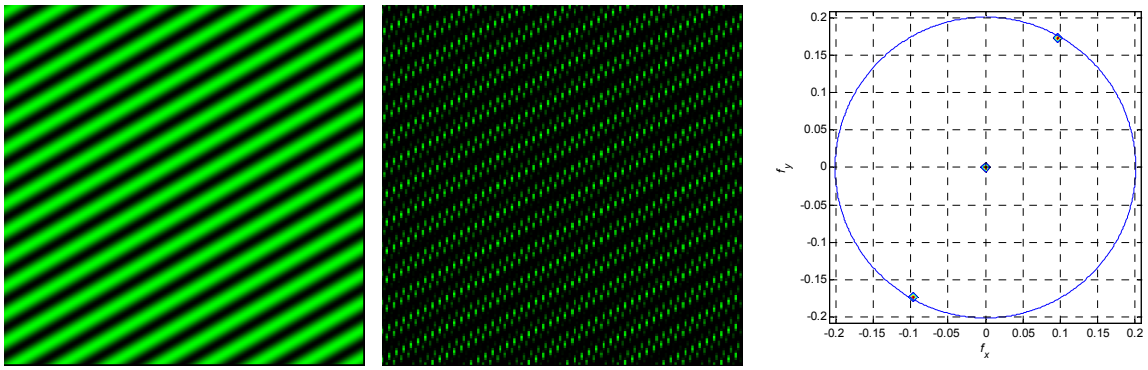
In the proposed method all analysis is done in the frequency domain. The method can be summarized by the following five steps. First, we prepare an image having a signal of a given frequency (f_{x_0}, f_{y_0}) with f_{x_0} and f_{y_0} being the frequencies of the input signal in horizontal and vertical direction, respectively. In this paper, all frequencies are normalized to one, with one corresponding to half of the sampling rate. Second, by applying the interleaving pattern and the visibility mask we prepare a simulated image (see Section 2.2). Third, we calculate the spectrum of the simulated image. We normalize the spectrum towards the input signal, that is, the amplitude of the input frequency component is normalized to be zero dB in the evaluated spectrum. Instead of having only peaks at the input signal frequency, due to the interleaving pattern, the spectrum of the simulated image has peaks on multiple places. We have found out that from the visual quality viewpoint, we are only interested in fequency components that have a lower frequency than the input signal [12].This corresponds to frequencies inside a circle with radius

$$r_0 = \sqrt{f_{x_0}^2 + f_{y_0}^2}.$$

Fourth, we apply a threshold criterion on the spectrum. The threshold level depends on the desired distortion. A lower threshold will correspond to tougher requirements on the image, and consequently, on a lower visible image distortion on the display. Fifth, after the threshold is applied, we check if there are any sinusoidal components left inside the circle with radius r_0 beside the DC component. If no, then we conclude that a signal with this particular frequency will be properly represented (visible) on the display. We define that this frequency is in the passband of the display. This is illustrated in Figure 4. As seen in Figure 4c, there are no frequency components above the threshold in the area of interest. Therefore, in Figure 4b we can see the most important features of the input signal. If there are signals present inside the circle with radius r_0, then the image will be represented on the display with distortion higher than desired. Therefore we do not want signals of such frequency in our input signal, that is, we have to filter it out. We will define that this frequency is in the stopband of the display. An example is illustrated in Figure 5. As seen in Figure 5c, there are peaks inside the radius of interest. Therefore, in Figure 5b the input signal is lost

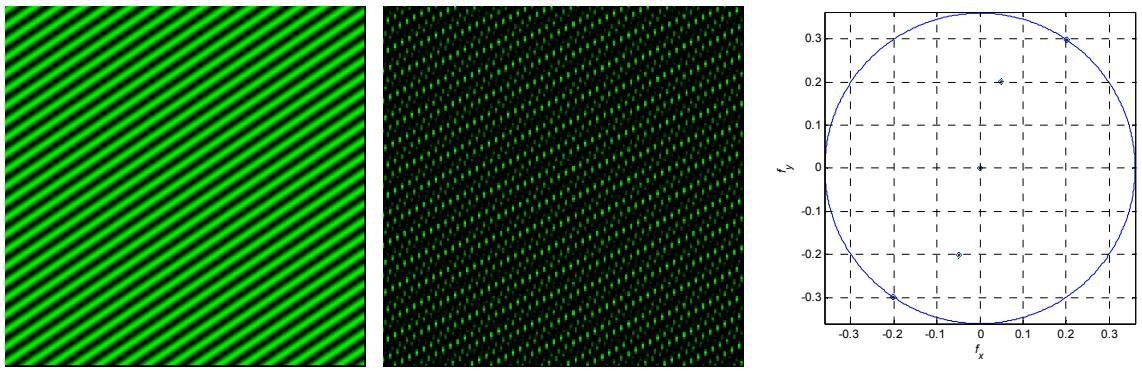
By repeating the above described procedure for images with signals of various frequencies in the intervals $f_x \in [-0, 1]$ and $f_y \in [-1, 1]$, we can determine all frequencies that will be properly represented on the screen, that is, the passband of the display. In the case of the X3D-display we estimate the passband as shown in Figure 6. In the figure, the simulated passband frequencies are marked with dots and the passband edge is emphasized with the blue line. Please note that the passband area presented in this figure (same applies to figures in the following sections) is discrete only due to our discrete simulation. We assume that the passband is continuous and covers all frequencies in the area bordered by the blue line. This figure tells us which frequency can be present in the input image on order for the image to be properly

represented on the screen. In this figure, a threshold of 26dB has been used. As described in [10], this threshold has been derived experimentally and it has turned out to be a good choice in practice.



a) b) c)

Figure 4, data analysis for signal $f_x = 0.1$ and $f_y = 0.175$ (signal in the passband): a) input signal, b) simulated output, c) spectrum in the area of interest.



a) b) c)

Figure 5, data analysis for signal $f_x = 0.2$ and $f_y = 0.3$ (signal in the stopband): a) input signal, b) simulated output, c) spectrum in the area of interest.

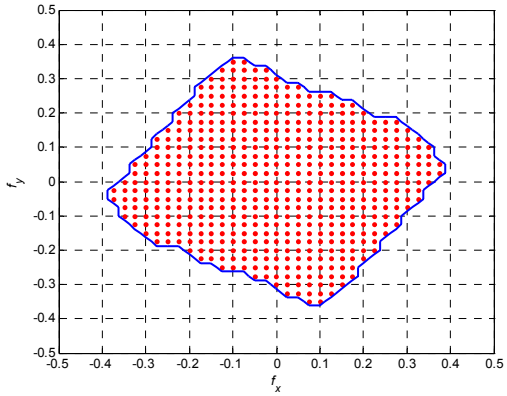
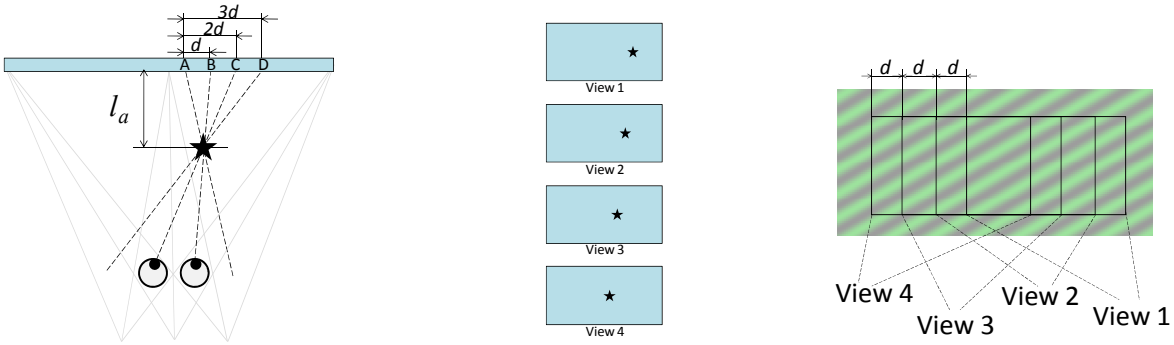


Figure 6, passband for the X3D-display evaluated based on the simulated data for 26dB threshold .

3. FREQUENCY PERFORMANCE VERSUS APPARENT DEPTH

3.1 Rendering objects with apparent depth

In a 3D scene, visualized on a multiview display, all objects are actually projected onto the surface of the display. However, since such display can visualize different images for a number of observation angles, it is possible to create false parallax and illusion of an object appearing in front of, or behind the display surface. Consider the example on Figure 7a. If a real object (e.g. the star-shaped mark on the Figure) appears in front of the display, according to the observation angle its projections appear on different position on the display (as marked with A, B, C, and D on the same Figure). If the multiview display visualizes a scene where the star-shaped object changes its position on the display as a function of the observation angle, this creates the impression of a virtual star-shaped object, hovering at apparent distance l_a in front the display, as exemplified in Figure 7a. This allows for a limited head parallax, where the observer can look at the scene from different angles. Furthermore, as each eye of the observer sees different view, the parallax of the virtual object creates stereoscopic illusion. Since eyes usually appear on horizontal line, and the distance between eyes is constant, the projections of the virtual object appear on equal distances on the screen surface as marked with A, B, C, D in Figure 7a. This corresponds to the constant disparity d in pixels between the observations of the same object in different views. For negative d (horizontal coordinate decreases with view number, as shown in Figure 7b) the object appears behind the surface of the display. For positive d the object appears in front of the display.



a) b) c)

Figure 7, disparity versus apparent depth: a) projections of a virtual object on the surface of a multiview display, b) position of a virtual object in different views, and c) cropping windows used for creating disparity between views.

3.2 Preparing test images

In order to study the display passband for test signals with different apparent depth, we rendered 2D test signals with different apparent depth. As input, we used test signals with various (known) frequencies (f_x, f_y). From each test signal we prepared a number of interleaved images with different apparent depth following a two step procedure. First, we prepared 8 views from each test image, by cropping the test image at different places as shown in Figure 7c. The cropping window for each view is shifted horizontally with an offset $s_n = d \cdot n$, where s_n is the offset for the n-th view, n is the view number and d is the targeted disparity. Then we interleaved the views into one interleaved test image. By changing the disparity d we simulate the process of putting the test signals at different apparent depths. In our experiments d varies between -10 and 10. To each interleaved test image, we applied the weighting mask which simulates the effect of the optical layer over the visibility of TFT elements. The weighting mask we used is the one derived for the X3D-display, and shown in Figure 2c.

3.3 Performance analysis for varying disparity

We applied the procedure described in Section 2.3 on all simulated images (images with various input frequencies and disparities) generated as described in the previous section. We used the same threshold criteria as earlier, namely, 26dB. In this way we estimated the frequency domain behavior of the display for different disparities. As an example, in Figure 8a and Figure 8b we show the display passband for the X3D-display for disparities 5 and 10, respectively. Based on these figures and Figure 6 it is obvious that different filters are needed for different disparity levels because the display passband has a different shape for different disparities.

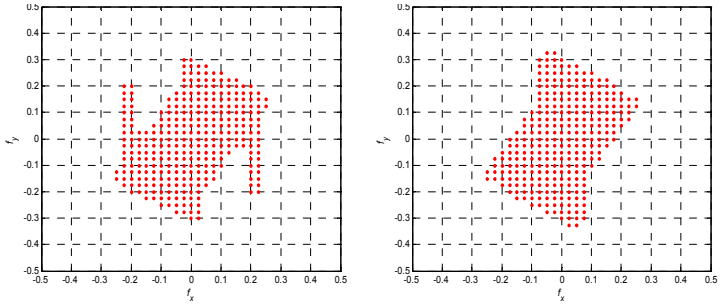


Figure 8, passband of the X3D-display based on simulated data for 26dB threshold and disparity: a) $d = 5$, b) $d = 10$.

4. FREQUENCY PERFORMANCE VERSUS PERCEIVED CROSSTALK

4.1 Perception of crosstalk

Ghosting artifacts are a form of image distortion. An observer labels some of distortion as being ghosting artifacts, if he or she is able to recognize repetitive structures and double contours. However, the human visual system (HVS) is not especially sensitive to ghosting artifacts in comparison to other structural distortions. HVS is optimized for perceiving the structure of the image, and is less sensitive to global contrast or brightness variance [11]. Many visual quality metrics attempt to assess the perceptual difference by estimating structural distortions of the image [12, 13]. The Weber-Fletcher law states that perceptibility of a change in stimuli is proportional to the amplitude of the stimuli. The works that assess visibility of crosstalk in stereoscopic images in typical observation conditions also measure the crosstalk as percentage of the input signal [5]. According to [5] the acceptable crosstalk varies between 5% (barely visible) and 30% (barely acceptable). We assume that visibility of image distortion is proportional to the visibility of crosstalk. Therefore, image distortion of 5% would be barely visible, and 30% distortion would be barely acceptable.

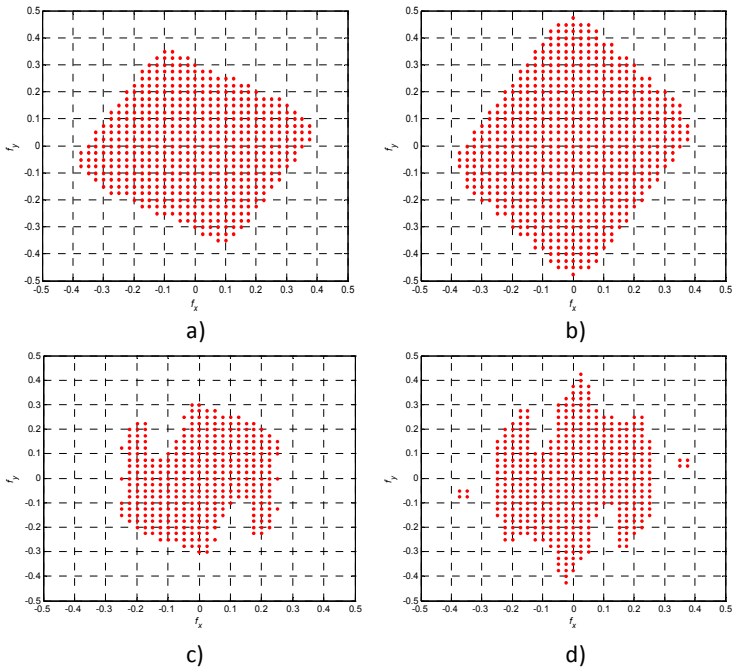


Figure 9, passband of the X3D-display based on simulated data: a) $d = 0; t = 10\%$, b) $d = 0; t = 30\%$, c) $d = 10; t = 10\%$, d) $d = 10; t = 30\%$.

4.2 Performance analysis for varying distortions

The perceived distortion values discussed in the previous section are defined in the spatial domain. Since we perform all processing in the spectral domain, we will transfer the results of that discussion to the spectral domain. For this purpose, we will assume that a distortion of t percentages in the spatial domain corresponds to t_{db} difference in decibels between the input signal and the strongest component inside the circle of interest. The relation between t and t_{db} is defined by the following well known expression:

$$t_{db} = -20\log_{10}(t/100).$$

In the case of the display under consideration, namely the X3D-display, the passband regions for two different disparities (0 and 10) and two different distortion values (10% and 30%) are given in Figure 9. This figure nicely illustrates that different filters are required for different values of distortion. Moreover, it confirms the observation from Section 3.3 that different filters are also required for different disparity levels.

Another important observation to be made here is related to the threshold of 26dB used in Sections 2.3 and 3.3. This threshold has been experimentally selected in [10] such that there are no visible (barely visible) distortions. According to the discussion in this section, 26dB corresponds to a 5% distortion. According to [5], 5% distortion is not visible. Since 26dB is also not visible (barely visible), this is an indirect proof that our assumption in this section related to transferring the results presented in [5] from spatial to spectral domain is valid.

5. ARTEFACT MITIGATION FRAMEWORK WITH SHARPNESS CONTROL

In order to visually optimize video content for a given multiview display, we propose artifact mitigation framework. It allows the user to specify the percentage of visible distortion over the original signal. The framework does the necessary processing to maintain the distortions within the selected limit, taking into account the display passband for different disparity values. It consists of two modules, shown in Figure 10 – offline processing module, where the display is measured and real-time processing module, which filters the input image according to its apparent depth and selected distortion limits. During the measurements in the offline processing module, one derives the passband of the display for a range of disparity values as explained earlier. Each passband can be approximated by a rectangle, for example as described in section 6.1 below. The output from the module is stored in two tables. One table contains the height of the equivalent passband for various disparity values and levels of distortion, and the other table – the corresponding width of the passband.

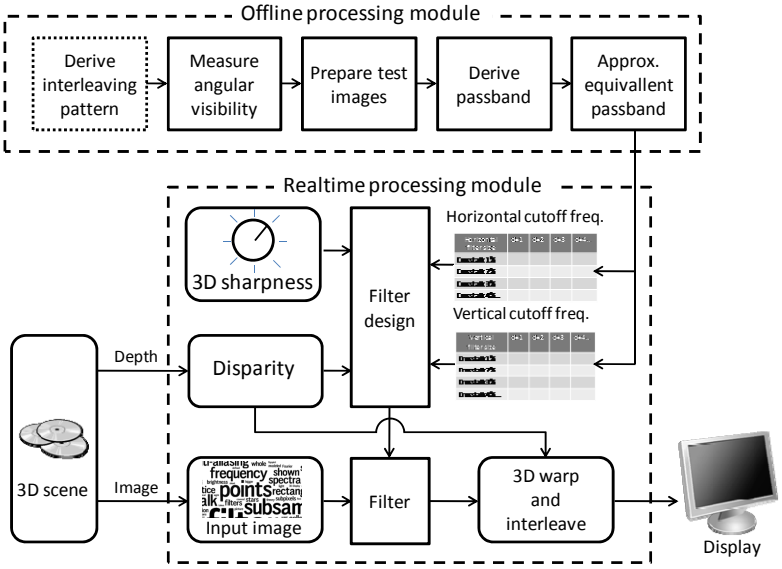


Figure 10, artifact mitigation framework.

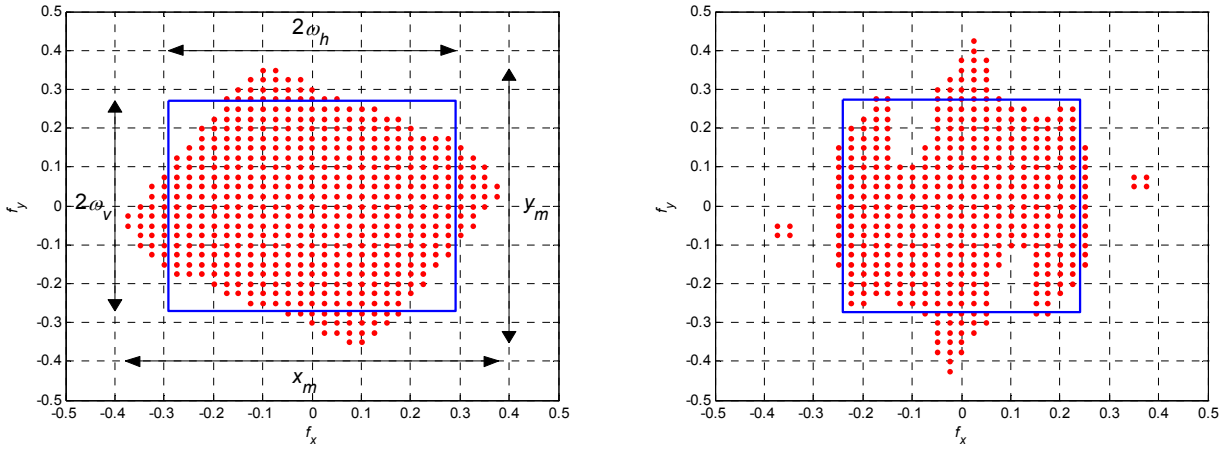
The realtime processing module uses these two tables to design the optimal filter for the input image. We consider that the input to the framework is a 3D scene in image-plus-depth format. Other possible inputs are subtitle track or GUI

widgets with known depth. The apparent depth encoded in the scene is then converted to suitable disparity value, according to the display size and resolution. The disparity value is used to select the corresponding column in each passband table. The user can set the value of the desired distortion level. We refer to this parameter as "3D-sharpness", since it controls the tradeoff between visibilities of details versus visibility of Moiré artifacts. The value of "3D-sharpness" is used to select the corresponding row of each table. The row and column selection in each table selects a cell. The values in the selected cells give the desired vertical and horizontal cutoff frequencies of an anti-aliasing filter. These cutoff frequencies are used for designing the filters. We describe one way to design such filter in Section 6.2 below. The filter is then applied on the input image before 3D warping and interleaving. Such filter mitigates the aliasing artifacts for the given disparity level and provides a desired level of "3D-sharpness".

6. DESIGN OF TUNEABLE ANTI-ALIASING FILTERS

6.1 Approximation of equivalent passband

The passband of a multiviev display for a given disparity and desired "3D-sharpness" has a nonuniform 2D shape (e.g. see Figure 9). In order to represent an image on the display properly we have to pre-filter the image with a filter having such passband as illustrated in Figure 10. In theory, we could design a 2D filter approximating the desired 2D shape. However, in practice, designing 2D filters is considerably more complicated than designing 1D filters and the implementation of a 2D filter can be computationally demanding. Therefore, in many cases, the desired 2D filter is approximated by two 1D filters, one for the horizontal direction and the other one for the vertical direction. Although, such 1D filters are just a rough approximation of the desired 2D shape, in turns out that in most cases this is good enough. For example, in our previous work [6, 7], we have shown that visually good results can be achieved by using separable 1D filters that approximate the required 2D shape. Another reason why 1D filters are good enough lies in the fact that in this paper we introduce the concept of user-tunable filters. Since the user can and will change the bandwidths of both filters according to his need, it is not too important to have 'perfect' filters.



(a) (b)

Figure 11, approximating the desired passband with a rectangle: a) $d = 0; t = 5\%$, b) $d = 10; t = 30\%$.

In order to use separable 1D filters, we have to approximate the desired passband shape with a rectangle. There are many ways how an arbitrary shape can be approximated with a rectangle. In our case, we will impose following two constraints: First, we want that the area covered by the rectangle is equal in size to the original passband area, and, second, the ratio between the maximum passband value in horizontal and vertical direction should be preserved. Taking these into account, it turns out that the parameters of the rectangle can be evaluated as:

$$\omega_h = \frac{1}{2}\sqrt{\frac{x_m}{y_m}A};\ \ \omega_v = \frac{1}{2}\sqrt{\frac{y_m}{x_m}A}\ ,$$

where $2\omega_h$ and $2\omega_v$ are the width and height of the rectangle, respectively, y_m and x_m are the maximum values of the passband in horizontal and vertical direction, respectively, and A is the passband area. These parameters are illustrated in

Figure 11a. Moreover, Figure 11a {Figure 11b} shows the approximations of passband for the X3D-display for disparity 0 and 5% distortion {disparity 10 and 30% distortion}. As it can be seen, particularly in Figure 11b, approximation is not always the best one due to the weird shapes of the desired passband. Nevertheless, as we demonstrate in Section 7, our approximation is good enough in practice. Parameters, ω_h and ω_v derived in the above manner correspond to the normalized cutoff frequencies of the two separable 1D filters.

By performing the approximation for various distortion levels and disparities, we end up with a set of cutoff frequencies in horizontal and vertical direction. These frequencies can be stored in a table, in order to be used by the realtime processing module as illustrated in Figure 10. It should be pointed out that although in practice, ω_h and ω_v, are evaluated for discrete values of disparity and distortion in order to have tables of reasonable size, we can easily interpolate for missing values. Linear interpolation has turned out to be satisfactory.

6.2 Filter design

For a given disparity and distortion level, based on the desired cutoff frequencies that are derived in a manner described in the previous section, we design two 1D filters. For designing each of these filters we used the windowing technique [14]. The used window was the Kaiser window with $\beta = 2.2$. This will result in filters with the first side lobe at approximately -30dB. This attenuation has turned out to be high enough in practice (similar conclusion like for distortion can be drawn, that is, -30dB corresponds to approximately 3% distortion). The windowing technique has been selected due to its simplicity. Since the idea is to design filters in real time based on the user preference and 3D content we need a fast design method.

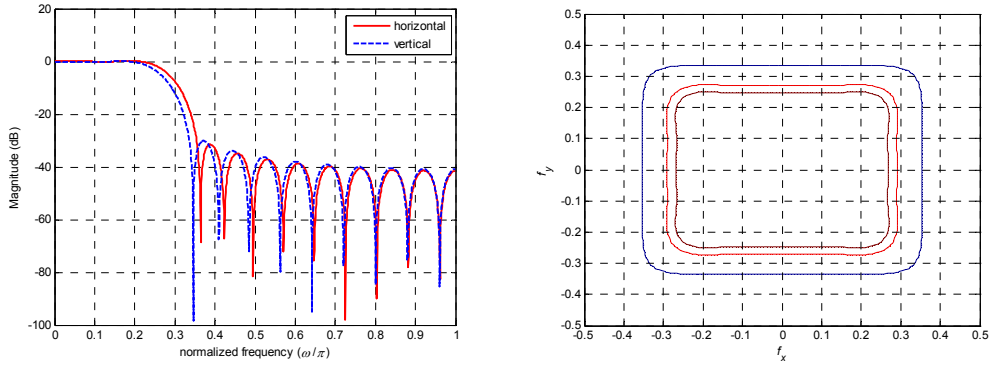


Figure 12, filters designed for the X3D-display for $d = 0; t = 5\%$: a) magnitude response of the vertical and horizontal filter, b) magnitude response of the corresponding 2D filter – contour for -3dB (innermost line), -6dB, and -30dB (outermost line).

As an example, the design of filters for the X3D-display for disparity 0 and 5% distortion is considered. The corresponding cutoff frequencies are $\omega_h = 0.291$ and $\omega_v = 0.272$. The magnitude responses of the designed 1D filters are given in Figure 12a. The magnitude (contour) of the corresponding 2D filter is given in Figure 12b. The -6dB level corresponds to the desired cutoff frequencies. The order of both filters is $N = 24$.

7. RESULTS

In Figure 13 we show the approximated X3D-display passband for several disparities between 0 and 10. Notably, the passband area does decrease monotonically with the increase of the disparity. The outer, red contours represent the passband area if distortion levels of 30% are allowed. The inner, blue contours represent the more strict, smaller passband area where distortion levels of less than 5% are desirable. One could see, that while the passbands for 30% distortion are always bigger than those for 5%, there is no visible relation between the two. Note, that for multiview displays with different interleaving pattern and visibility mask the relation between passband area and disparity of the image will most probably change.

Some examples of the effect of the designed anti-aliasing filters can be seen in Figure 14. All images on that figure are simulated output of the display as it would be seen directly from the front. The top row of images (Figure 14a-c) are rendered with zero disparity, while the bottom row (Figure 14d-f) are rendered with disparity $d = 5$. Images in the left

column are simulated with no pre-filter, and the color banding and Moiré artifacts are clearly visible. Images in the center column are prefiltered with the anti-aliasing filter for $t = 30\%$ and the corresponding disparity of $d = 0$ (top) and $d = 5$ (bottom). One can see that most artifacts are mitigated, but some residual aliasing is visible. The images in the right column are pre-filtered with the anti-aliasing filter for $t = 5\%$, and exhibit even less artifacts, however at the expense of texture loss.

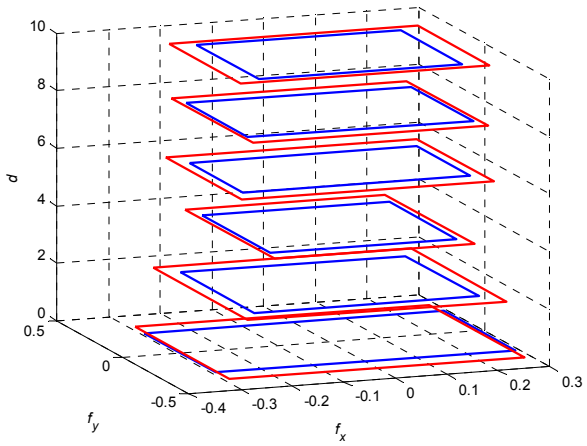


Figure 13. passband for disparities $d = 0,2,4,6,8,10$, for $t = 5\%$ (blue), and $t = 30\%$ (red).

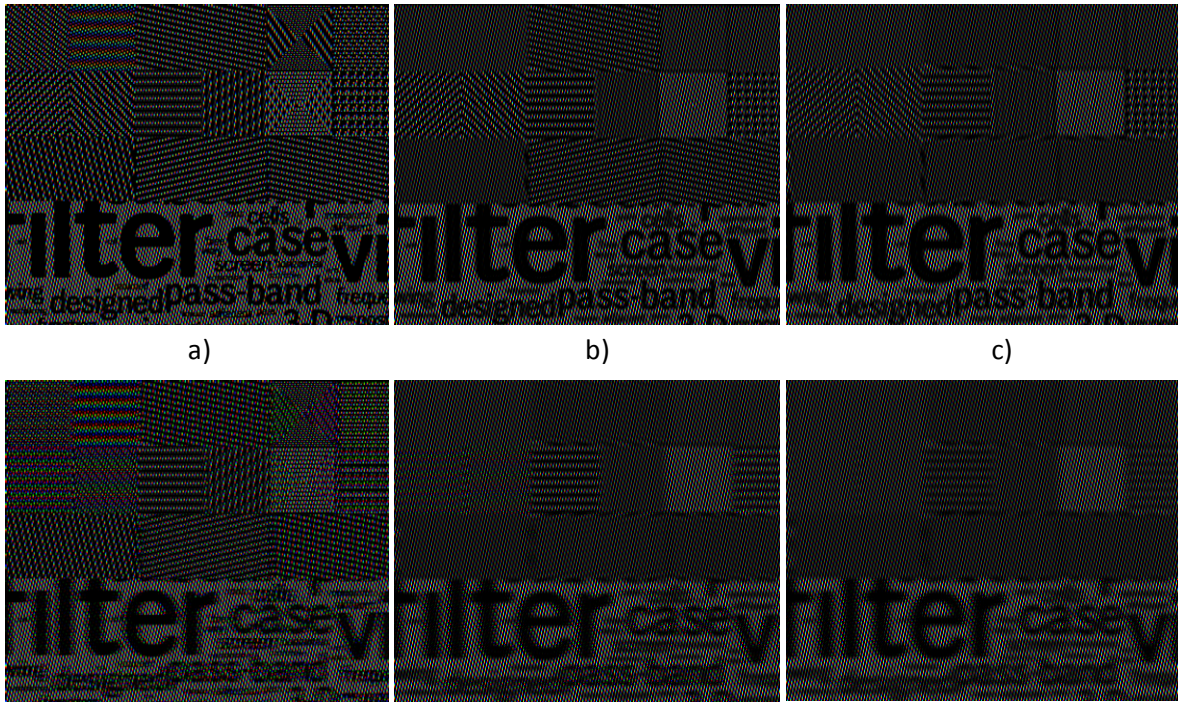


Figure 14. effect of the designed anti-aliasing filters for different disparity and distortion levels (simulated outputs): (a-c) test image with no disparity, a) without pre-filtering, b) pre-filtered with the filter for $d = 0; t = 30\%$,c) pre-filtered with the filter for $d = 0; t = 5\%$ (d-e) test image with disparity 5, d) without pre-filtering, e) prefiltered with the filter for $d = 5; t = 30\%$, f) pre-filtered with the filter for $d = 5; t = 5\%$.

8. CONCLUSIONS

We proposed a model of a multiview display which considers the display as an image processing channel. The model assumes that the input is in image-plus-depth format. We used measurement data to construct the visibility mask of the display, and used it to simulate the output of that channel. We analyzed the distortions, introduced by the channel for test images with various frequency components and disparity values. We derived multiple so-called passbands, which define combinations of frequency components and disparity values, which pass through the channel with distortions lower than the given threshold. We proposed methodology for design of tunable filters, which can be used for realtime anti-aliasing of multiview 3D images. Such tunable filters allow the user to select the desired level of "3D-sharpness" and control the trade-off between visibility of details and that of artifacts. Finally, we gave some practical results for images, filtered with anti-aliasing filters optimized for one 8-view autosterescopic display.

9. ACKNOWLEDGMENTS

The work is partially sponsored by Academy of Finland (project no. 213462, Finnish Programme for Centres of Excellence in Research 2006-2011) and Nokia Scholarship Grant provided by the Nokia Foundation.

REFERENCES

1 S. Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

2 N. Dodgson, "Autostereoscopic 3D Displays," *Computer* , vol.38, no.8, pp. 31- 36, Aug. 2005, IEEE, 2005.

3 W. IJsselsteijn, P. Seuntiens and L. Meesters, "Human factors of 3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

4 J. Konrad and P. Agniel**,** "Subsampling models and anti-alias filters for 3-D automultiscopic displays," *IEEE Trans. Image Process.*, vol. 15, pp. 128-140, Jan. 2006.

5 F.Kooi, A. Toet, "Visual comfort of binocular and 3D displays", Displays, Volume 25, Issues 2-3, August 2004, Pages 99-108, ISSN 0141-9382, DOI: 10.1016/j.displa.2004.07.004, 2004.

6 A. Boev, R. Bregovic, A. Gotchev, "Methodology for design of anti-aliasing filters for autostereoscopic displays", Special issue on Advanced Techniques on Multirate Signal Processing for Digital Information Processing*, Journal of IET Signal Processing*, to be published (2010).

7 A. Boev, R. Bregovic, A. Gotchev, K. Egiazarian, "Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays", in Proc. *of The 2009 International Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2009*, Helsinki, Finland, 2009.

8 X3D-23" Users' Manual. NewSight GmbH. Firmensitz Carl-Pulfrich-Str. 1 07745 Jena, 2006.

9 A. Nashel and H. Fuchs. "Random hole display: A non-uniform barrier autostereoscopic Display". In proc. of *3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video*, 2009.

10 A. Boev, R. Bregovic, A. Gotchev, "Measuring and modeling per-element angular visibility in multiview displays", Special issue on 3D displays, Journal of Society for Information Display, 2010.

11 B. Wandell, "*Foundations of Vision*", Sinauer Associates, 1995.

12 Z. Wang, A. Bovik, H. Sheikh and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Processing*, vol. 13, No. 4, pp. 600-612, 2004.

13 S. Winkler, "Perceptual Video Quality Metrics – A review", in H. Wu and K. Rao, eds. "*Digital video image quality and coding*", ch. 5, CRC press, 2006.

14 S. K. Mitra, *Digital signal processing: A computer based approach*, New York: McGrow-Hill, 2005.

[P07] A. Boev, R. Bregovic, D. Damyanov, A. Gotchev, “Anti-aliasing filtering of 2D images for multi-view auto-stereoscopic displays”, in Proc. of *The 2009 International Workshop on Local and Non-Local Approximation in Image Processing, LNLA 2009*, Helsinki, Finland, 2009

This page left blank

ANTI-ALIASING FILTERING OF 2D IMAGES FOR MULTI-VIEW AUTO-STEREOSCOPIC DISPLAYS

Atanas Boev[1], *Robert Bregovic*[1], *Damyan Damyanov*[2], *and Atanas Gotchev*[1]

[1] Department of Signal Processing, Tampere University of Technology, Finland,
[2] Department of Telecommunications, Technical University of Sofia, Bulgaria
[1] firstname.lastname@tut.fi, [2] ellov@abv.bg

ABSTRACT

Abstract – In this paper, we address the problem of anti-aliasing filtering of images to be displayed on auto-stereoscopic displays. Auto-stereoscopic displays are constructed to create 3D visual effect by no special glasses but utilizing extra optical layer to cast different images to different directions. The topology of such layer is a compromise between the number of different views generated and the spatial resolution per view being fraction of the full 2D spatial resolution. Usually, the compromise is achieved by slanted and non-rectangular sub-sampling grids causing however corresponding aliasing artefacts. These artefacts are especially visible and annoying when 2D imagery, such as graphics and text, is to be displayed on auto-stereoscopic displays. In our work, we design efficient anti-aliasing filters to mitigate this effect. Two classes of filters are studied for a 3D display case. The first class is the class of non-separable filters, which takes into account the non-rectangular topology of the particular sub-sampling grid and the effect of inter-view crosstalk, and aims at suppressing the respective aliasing replicas appearing in non-rectangular positions on the 2D Fourier plane. The second class is the class of efficient separable 2D filters based on 1D anti-aliasing filter design. We demonstrate that the latter class results in subjectively higher quality images. Studying this particular case further, we design filters for different types of imagery, distinguishing between text and graphics and also between 'smooth' and 'sharp' target anti-aliased images. As it is difficult to quantify the results by objective measures, we illustrate them by visual examples. Subjective inspections have also confirmed the feasibility of our approach.

1. INTRODUCTION

3D displays aim at delivering the perception of depth (the third dimension). Certain types of 3D displays recreate 3D scenes without requiring the observer to wear special glasses. Such 3D displays are known as auto-stereoscopic displays, and they work by casting two or more different images each one visible from different angle. Due to this principle of operation, only a subset of all image pixels is visible from a particular angle. The visible pixels appear on a non-rectangular grid, and rendering images on this grid requires special anti-aliasing filters [1], [2].

A 3D display may be used to visualise a combination of 2D and 3D objects, or 2D content only, if 3D content is not available. In a mixed scene, aliasing artefacts are especially visible in 2D objects [1], [4]. Our work studies two sets of filters, which can be used for anti-aliasing of such content.

The paper is organized as follows: the next section briefly explains how multi-view displays work, and how its optical characteristics can be measured and modelled. Results for a particular multi-view display – namely 23" X3D produced by NewSight – are given as example. Sections 3 and 4 present two different approaches in designing anti-aliasing filters optimized for a specific 3D display. In Section 3, optical measurements of grid topology and interview crosstalk are used for designing non-separable 2D anti-aliasing filter for the 23" X3D display, while Section 4 presents an attempt to reproduce the same results using separable filters. Different filters are proposed for "image" and "text only" 2D content.

Finally, the visual quality and computational intensity of different anti-aliasing filters is compared in Section 5. Both simulated images and snapshots of 23" X3D display showing different test images are given as example.

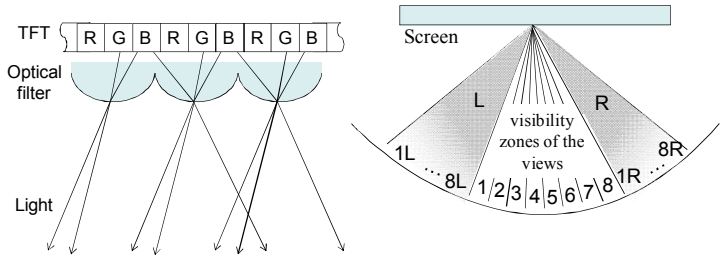
2. MULTIVIEW DISPLAYS

2.1. Principles of operation

Modern multi-view displays use TFT matrix for image generation [1], [5], [6]. An optical filter is mounted on the surface of the display as shown in Figure 1a. The filter redistributes the light coming from the TFT towards different horizontal directions.

The set of sub-pixels, visible from given direction form a colour image, also known as a *view*. The range of angles, from which a view can be seen, is known as the *visibility zone* of that view. Usually, the visibility zones of all views appear in horizontal direction in front of the display, as depicted in Figure 1b.

In order to visualize a scene in 3D, a number of different observations of that scene should be simultaneously shown on the 3D display. The process of mapping an image to the sub-pixels corresponding to one view is called *view interleaving* or *view interdigitation* [1]. The map of correspondences between addressable sub-pixels of the display and the view they belong to is called *interdigitation map*. Usually the interdigitation map has repetitive structure, which can be represented by an *interdigitation pattern* copied multiple times over the display surface.



a) b)

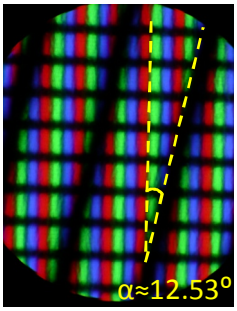
Figure 1. Operation principles of a multi-view display: a) optical filter (view from the top) and b) visibility zones of the views (view from the top)

When 3D object is visualized on a multi-view display with n views, n different observations are interleaved into one compound 3D image. A 2D object, which is not meant to appear floating in front or behind the screen surface, is represented by n identical observations. In this case, the optical filter can be regarded as a mask, which partially covers the underlying 2D image.

2.2. NewSight 23" X3D multi-view display

The multi-view display studied in this paper is *23" 3D-Display AD* built by X3D-Technologies GmbH. The display uses 23" TFT monitor with resolution of 1920x1200. The display area is 495x310mm and the optimal distance for observing the 3D effect is 1.5m. The X3D display is marketed as 8 view 3D display with interdigitation pattern as shown in Figure 2a [7].

	Column 1			Column 2			Column 3		
row	r	g	b	r	g	b	r	g	b
1	1	2	3	4	5	6	7	8	1
2	2	3	4	5	6	7	8	1	2
3	2	3	4	5	6	7	8	1	2
4	3	4	5	6	7	8	1	2	3
5	4	5	6	7	8	1	2	3	4
6	4	5	6	7	8	1	2	3	4
7	5	6	7	8	1	2	3	4	5
8	6	7	8	1	2	3	4	5	6
9	6	7	8	1	2	3	4	5	6
10	7	8	1	2	3	4	5	6	7
11	8	1	2	3	4	5	6	7	8
12	8	1	2	3	4	5	6	7	8



a) b)

Figure 2. X3D 23" display: a) interdigitation pattern, b) micrograph of the optical filter

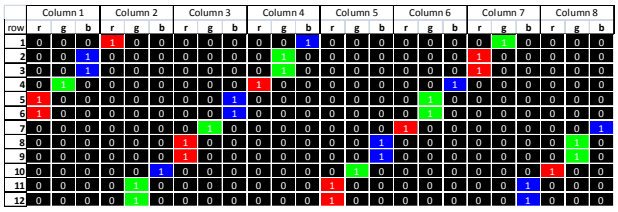
The optical filter is called *wavelength-selective filter array* [5], which works similarly to a parallax barrier – blocking the light in some directions and passing light in others. The filter array has regular structure, mounted on the display at a slant of 12.53 degrees, as can be seen from the micrograph in Figure 2b.

Following the interdigitation pattern, one can construct a map of sub-pixels visible from certain angle – for example the sub-pixels marked with "1" in Figure 3a. For different observation angles the map of visible sub-pixels is the same, only shifted in horizontal and vertical direction.

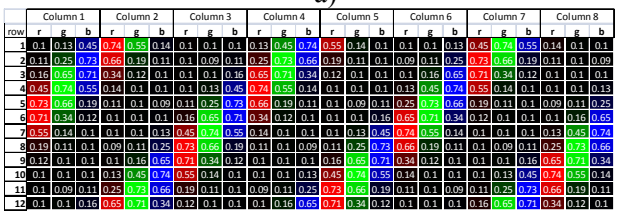
2.3. Crosstalk

Multi-view displays suffer from two common artefacts - *image flipping*, caused by the noticeable transition between the viewing zones, and *picket fence effect*, caused by optical filter magnifying the gaps between sub-pixels. The common practice to mitigate this effect is to broaden the observation angle of each view, thus interspersing the visibility zones [5]. As a result, from a particular angle, a number of views are simultaneously seen. The view originally intended to be seen is the brightest one, but its neighbouring view are visible as well. This effect can be regarded as inter-channel crosstalk.

Methodology for crosstalk estimation and measurement results for 23" X3D were presented in [4]. Based on these measurements the optical masking pattern is reconstructed as shown in Figure 3b. The masking pattern defines the set of sub-pixels visible on the display from a particular direction, as well as their relative brightness in the range between 0 (black) to 1 (maximum brightness).



a)



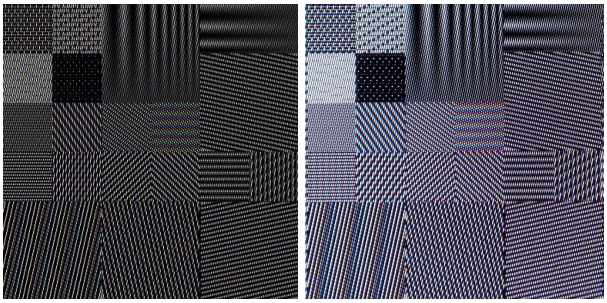
b)

Figure 3. Map of visible sub-pixels from a particular angle: a) without crosstalk, b) with crosstalk

2.4. Aliasing

Selective masking of a 2D image caused by optical filter can be modelled as a sub-sampling on a non-orthogonal grid. Without pre-filtering this process creates aliasing artefacts.

An example for aliasing artefacts on 23" X3D display is given in Figure 4, where simulated image is shown next to an actual photograph of the display. The original 2D image can be seen in Figure 20a. Aliasing process is simulated by using the sub-pixel visibility map from Figure 3a as a mask. The masked image, shown in Figure 4a exhibits noticeable aliasing artefacts. Alternatively, the 2D image was visualized on the 23" X3D display and a photo was taken. The photographed image is shown in Figure 4b.



a) b)

Figure 4. Aliasing, caused by optical filter: a) simulated image (fragment) and b) actual photograph of the X3D display (fragment). The original image used for the experiment is shown in Figure 20a

In multi-view displays aliasing artefacts appear in all types of scenes, but are especially visible in 2D content, as in 3D images aliasing is somewhat masked by more severe artefacts such as ghosting [1][4]. While most of content created for multi-view display would be in 3D, there are cases where 2D images would appear on such display as well. The typical cases include 2D graphics, 2D text and natural 2D images.

In order to eliminate aliasing errors due to the non-orthogonal sub-sampling pattern of the display, in this paper two different methods for designing anti-aliasing filters have been applied. The first one is based on non-separable filters and the second one on separable ones. These methods are discussed in detail in the following two sections.

3. NON-SEPARABLE ANTIALIASING FILTERS

First method for designing anti-aliasing filters is based on the method introduced by Jain and Konrad in [1] (in the rest of the paper it will be referred to as the JK method). JK method can be used for designing 2D non-separable anti-aliasing filters for an arbitrary sub-sampling pattern. It is assumed that a 2D image is processed and correspondingly, the method works best for 2D imagery.

The basic idea of the JK method is to design a 2D filter in such a way that the passband of the filter spans all frequencies at which the contribution of all alias terms is smaller than the original signal itself. The stop-band of the filter is assumed to span all other frequencies. This is achieved by the following steps: First, based on the sub-sampling pattern the position and intensity of all aliasing terms in the 2D frequency domain is estimated. Second, the contribution of these aliasing terms to the overall spectrum of a given image is evaluated. Instead of using a particular image, an image model is utilized. In [1], the use of a first-order 2D Markov model is suggested for modelling the image. Third, the passband of the ideal filter is selected as the region where the spectrum of the signal is greater than the contribution of all aliasing terms. The rational for this is that all frequencies should be preserved for which the signal is stronger than the aliasing terms. Finally, fourth, a filter design technique is applied to design the filter itself based on the above determined specifications.

In [1], the JK method has been applied for the 9-view SynthaGram SG202 monitor. In our work, we reproduce their approach for the case of the X3D display under consideration. We specifically utilize our measurements of the grid topology and the inter-view crosstalk obtained for that display. We consider two cases: crosstalk free case and crosstalk-aware case. The first case assumes that the viewer can see, while at an observing position corresponding to a certain view, pixels belonging to that view only. In the crosstalk-aware case, it is assumed that the viewer can see, in addition to the pixels from the principal view, also pixels from the adjacent views.

3.1. Crosstalk-free case

In this case the display is considered ideal, that is no interference between neighbours channels exist. For an ideal display with n views the viewer sees only $1/n$ pixels of the display, or, roughly speaking, about $1/n$ of the whole brightness. Depending on the sub-sampling patterns, in one view a non-uniformly sampled image is seen. Based on the known sub-sampling pattern it is possible to estimate which frequencies in the original image have to be suppressed in order to avoid aliasing.

For the X3D display, the sub-sampling patterns for all colors in all views differ only by a shift in the space domain. There is no difference in the spectral domain. Therefore, in this paper, without loss of generality,

when designing filters, the sub-sampling pattern of the red component for the fourth view has been used. This sub-sampling pattern is shown in Figure 5.

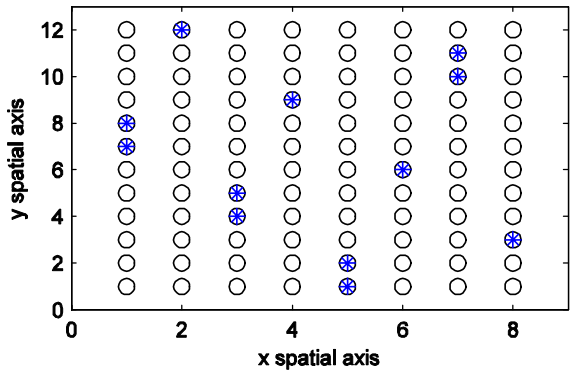


Figure 5. Sub-sampling pattern for the fourth view. Circles and stars represent the orthogonal grid and the sub-sampling pattern, respectively.

The JK method is focused on the non-orthogonal sub-sampling of the image for each view. Taking only one of n points of the image, where n is number of views ($n=8$ for the X3D display), a significant change in the spectrum of the picture occurs. Essentially, this change can be explained with the occurrence of attenuated replicas of the original spectrum on certain points of the 2D spectral domain. In the well known case of a sampled 1D signal, replications of the original spectrum occur on points, which are equal to the integer multipliers of the sampling frequency. In the 2D case with non-orthogonal sampling the replicas occur on various positions depending on the sub-sampling pattern. For the sub-sampling pattern given by Figure 5, the position and amplitude of the main spectral component and all replicas are shown in Figure 6. The main spectral component (baseband spectrum) is located at $(f_x,f_y)=(0,0)$. All other peaks represent positions of spectrum replicas.

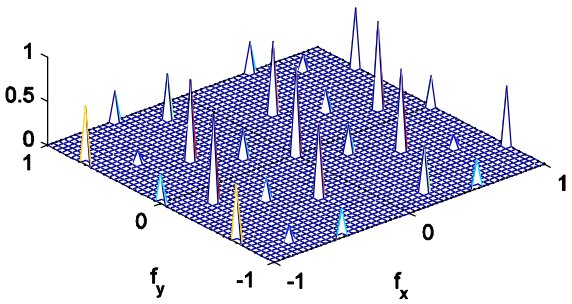


Figure 6. Spectrum of the sub-sampling pattern for the fourth view.

In order to determine the contribution of all alias terms independent of the image itself, a statistical representation of a real image has been used. As in [1], a 2D separable autocorrelation model $R_u[m,n]=\rho^{|m|}\rho^{|n|}$ has been utilized. Here, $0<\rho<1$ is the correlation coefficient and is typical chosen in the 0.9-0.99 range. This model is based on the first-order Markov model as discussed in [1], [9]. The power spectral density of this autocorrelation model is:

$$\Phi(f_x,f_y)=\frac{1}{\pi^2}\frac{f_0^2}{(f_x^2+f_0^2)(f_y^2+f_0^2)}, \qquad (1)$$

with $f_0=-(\ln\rho)/(2\pi)$ and f_x and f_y being the normalized frequencies. The power spectral density function for $\rho=0.9$ is shown in Figure 7.

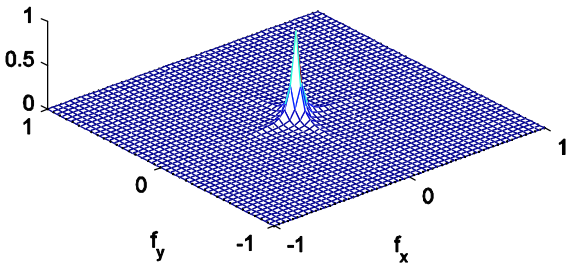


Figure 7. Power spectrum density function for the used autocorrelation model with $\rho = 0.9$.

In terms of classical anti-aliasing theory one should take the original spectra and try to suppress all other replicas in order to obtain an alias-free signal. In practice, this would be too restrictive. Therefore, in the JK method a different approach is used. The idea is to widen the pass-band of the anti-aliasing filter as much as possible. This can be done by defining the boundary of the passband as follows: The cut-off frequencies of the filter are set at points where the amplitude of the original spectrum equals to the sum of all amplitudes of the spectral replicas. This can be formulated as:

$$H_d(f_x, f_y) = \begin{cases} 1, if\ \Phi(f_x, f_y) > K\sum_i \beta_i \Phi(\xi_x^i - f_x, \xi_y^i - f_y) + \varepsilon \\ 0, \quad \text{otherwise} \end{cases} \qquad (2)$$

The variables (f_x, f_y) represent the spectral coordinates across x and y spectral axis, respectively. H_d is the desired frequency response of the anti-aliasing filter, (ξ_x^i, ξ_y^i) are the coordinates along x and y axis at which the i-th spectral replica occurs and β_i is its associated gain. The constant K permits changes in the pass-band of the filter. In this paper $K = 1$ has been used. Finally, ε is used to eliminate some regions in corners of the spectrum that due to the used image model could be declared as passbands. In this paper, for all designs, $\varepsilon = 0.001$ has been used.

By evaluating Equation (2), the desired frequency response of the ideal 2D filter $H_{id}(f_x, f_y)$ is shown in Figure 8. For designing a 2D filter approximating this ideal one, the widowing method with the Hamming window of length $N = 49$ has been used (for more detail see `fwind2` function in Matlab). The impulse response and the magnitude response of the designed filter are shown in Figure 9 and Figure 10, respectively. As seen on Figure 3a, only $1/8^{th}$ of the available pixels are visible in one view. Therefore it is to expect that the passband area of the filter should be approximately 1/8. However, the passband surface of the filter designed for the one view case is considerably smaller – 0.037. This proves that assuming crosstalk-free view model is too restrictive for the given visibility map. Therefore in the next section the contribution of adjacent views is taken into account.

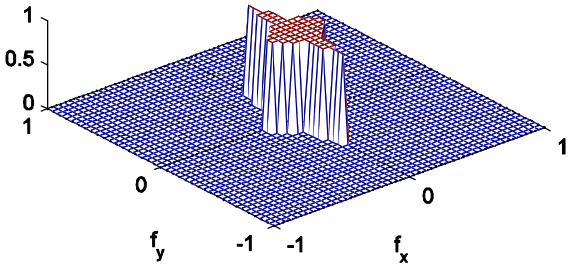


Figure 8. Magnitude response of the ideal filter in the one view case.

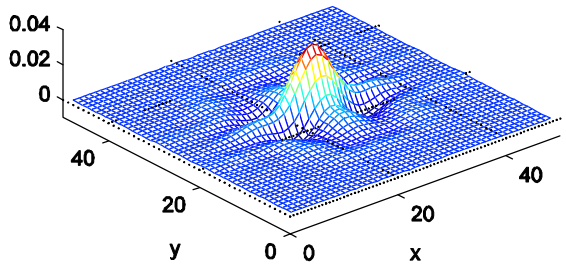


Figure 9. Impulse response of the filter in the one view case.

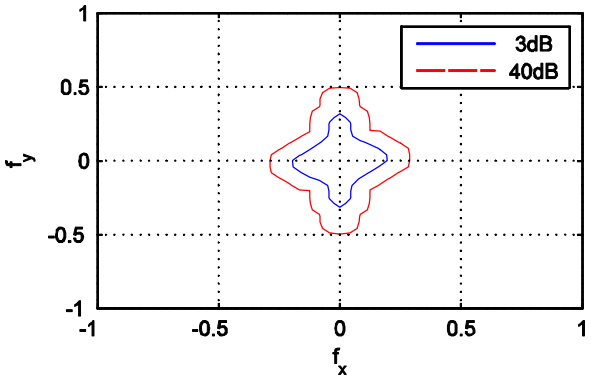


Figure 10. Magnitude response (contour) of the filter in the one view case.

3.2. Crosstalk-aware case

In the previous section, it was assumed that in one view only pixels from that view are seen. In practice, as it has been experimentally observed [1], [4], in addition to pixels from the view under consideration, pixels from adjacent views are also visible, although with a smaller intensity. For the display under consideration, the sub-sampling pattern for fourth view with contributions from the adjacent third and fifth view is shown in Figure 11. It has been measured that the contributions of those three views are, in average, as shown in Table I.

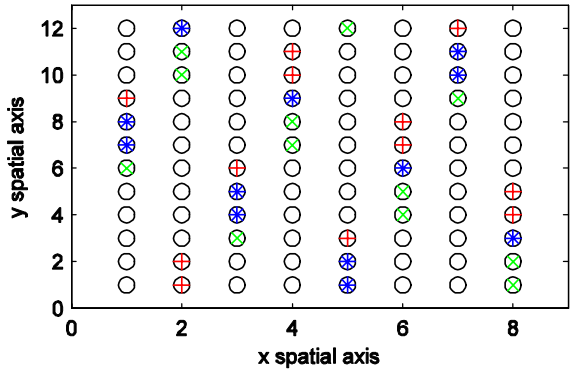


Figure 11. Sub-sampling pattern for the fourth view with contributions from the third and fifth view. Circles represent the orthogonal grid. Stars (blue), pluses (red) and x (green) are the sub-sampling patterns for fourth, third and fifth view, respectively.

Table I Contributions of individual views to the fourth view

View	#3	#4	#5
Amplitude	0.43	1.00	0.43

For this sub-sampling pattern, the spectrum is shown in Figure 12. Again, the term at $(f_x, f_y) = (0, 0)$ corresponds to the original, non-aliased term. It can be observed that the position of aliasing terms is the same as

for the one view case. However the contribution (amplitude) of each alias component is smaller than in the one view case. This can be interpreted that due to the crosstalk more of the image is seen in one view and as such the overall aliasing is smaller.

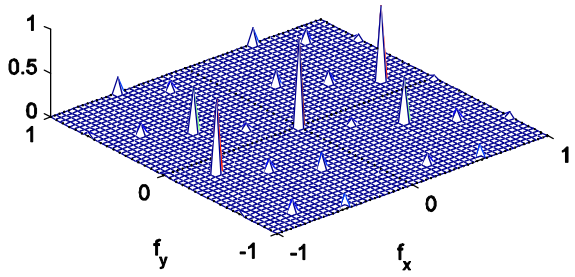


Figure 12. Spectrum of the sub-sampling pattern for the fourth view with contributions from third and fifth view.

By applying the same procedure as discussed in the previous section, specifications for the ideal filter can be determined. This ideal filter is shown in Figure 13. The impulse response and the magnitude response of the designed filter are shown in Figure 14 and Figure 15, respectively. The passband surface of the ideal filter is 0.068. This is still less than 1/8. It can be concluded that both crosstalk-free and crosstalk-aware filters are too restrictive (this will be also shown in the example section) and as such they are not performing well in practice. Moreover, they are computational heavy. Therefore in the next section a separable filter design is proposed that generated filters that are fast and perform well in practice.

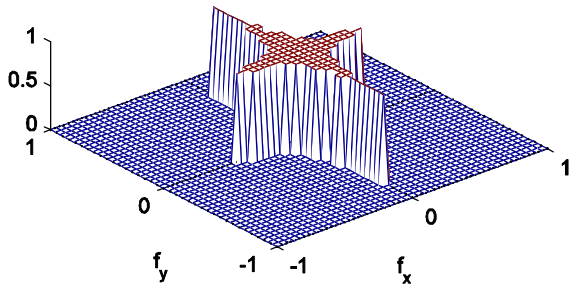


Figure 13. Magnitude response of the ideal filter in the crosstalk case.

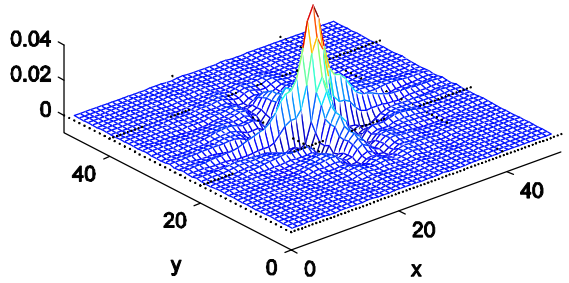


Figure 14. Impulse response of the filter in the crosstalk case.

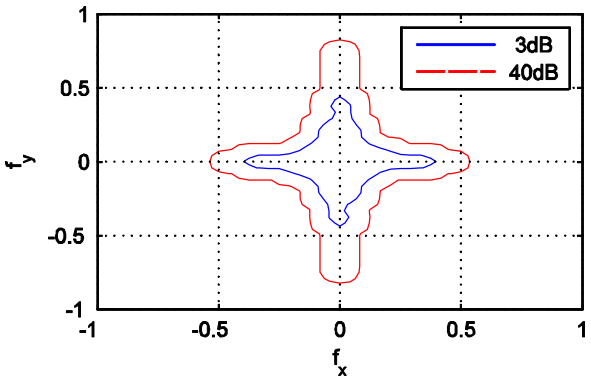


Figure 15. Magnitude response (contour) of the filter in the crosstalk case for attenuations 3 dB and 40 dB.

4. SEPARABLE ANTI-ALIASING FILTERS

In this section a method is proposed for designing separable filters for reducing aliasing errors when showing a 2D image on a 3D screen. The motivation of using separable filters is threefold. First, separable filters are, in principle, easier to design than non-separable 2D filters, second, they are computationally more efficient than non-separable filters of the same size, and, third, by properly designing these filters (allowing some alias errors in the image), visually (subjectively) better results are expected than by using the method described in the previous section.

When using separable filters to filter an image, the image is first filtered in horizontal direction and then, in some case, in vertical one. This will be referred to as the horizontal and vertical filtering. Parameters related to horizontal and vertical filters will be denoted by subscript *h* and *v*, respectively. Consequently, instead of one 2D filter, two 1D FIR filters have to be designed.

4.1. 1D filter design

There are many methods for designing FIR filters. When selecting a design method to be used in this paper, following two criteria were taken into account: First, the method has to be fast because many filters had to be designed in order to choose the best one for the given problem. Second, the filters should have enough attenuation in the stopband in order to suppress aliasing terms. A good candidate satisfying the above two conditions are FIR filters designed in the least-mean-square (LMS) sense [10]. In this filters the energy of the error in stopband and passband is minimized. The design problem for an FIR filter with transfer function

$$H(z) = \sum_{n=0}^{N} h[n] z^{-n} \qquad (3)$$

can be stated as follows: For a given filter order N, passband and stopband edges ω_p and ω_s minimize

$$E_2 = \int_0^{\omega_p} \left(H(e^{j\omega}) - e^{-j\omega N/2} \right)^2 d\omega + \int_{\omega_p}^{\pi} H(e^{j\omega})^2 \, d\omega . \qquad (4)$$

After some mathematical manipulations, the above problem can be rewritten as

$$E_2 = \boldsymbol{h}^T \boldsymbol{Q} \boldsymbol{h} + \boldsymbol{h} \boldsymbol{b} + c . \qquad (5)$$

Here, $\boldsymbol{h}$ is the vector containing the filter coefficients, and $\boldsymbol{Q}$ and $\boldsymbol{b}$ are a matrix and a vector that depend only on the filter order and passband and stopband edges and c is a constant [10]. The energy of the error is minimized for

$$\boldsymbol{h} = \boldsymbol{Q}^{-1} \boldsymbol{b} , \qquad (6)$$

that is, in order to design a filter for a given N, ω_p, and ω_s, only a system of linear equations has to be solved. Moreover, the energy in the stopband is minimized, which in turn, minimizes the aliasing error. Therefore, both of the initial criteria are taken care of.

Due to the non-orthogonal sampling pattern, it is not easy to decide what values for filter order, passband and stopband edges should be selected. Therefore, subjective experiments have been performed to determine good values of these parameters. The experiments have been carried out by using following steps:

Step 1. An appropriate image has been selected.

Step 2. The initial filter orders, passband edge and stopband edge were selected as $N_h = 30$, $\omega_{ph} = 0.8$, $\omega_{sh} = 0.9$, $N_v = 30$, $\omega_{pv} = 0.8$, $\omega_{sv} = 0.9$. The motivation of choosing these parameters lies in the fact that the difference between images filtered with such filters and non-filtered images, when seen on the X3D display, is negligible. As the goal is to improve the images after filtering, this turned out to be a good starting point.

Step 3. First edges ω_{ph} and ω_{sh} and then filter order N_h were reduced until an image of satisfied quality (described in more detail in following sections) is achieved. As this parameters influence filtering in the horizontal direction, only features containing vertical lines (and ones close to vertical) have be considered during this step. The minimal values for parameters N_h, ω_{ph}, and ω_{sh} are considered to be the optimal ones.

Step 4. Keeping the horizontal parameters determined in Step 3, repeat Step 3 by reducing parameters ω_{pv} and ω_{sv} and then filter order N_v. In this case, attention has been paid to horizontal lines and the ones close to them. Again, the minimal values for parameters N_v, ω_{pv}, and ω_{sv} are considered to be the optimal ones.

The above procedure has been repeated three times, twice for an image containing various patterns (patterns have been chosen in such a way to emphasize aliasing errors due to the sub-sampling patterns of the display) and once for text. This is described in the following sections.

4.2. Anti-aliasing of 2D images

For anti-aliasing of 2D images with separable filters, in this paper, two different approaches have been proposed. They will be refereed to as the "smooth" and "sharp" approach and are presented in the following two sections.

4.2.1. "Smooth" approach

In the "smooth" approach the goal was to design filters that will eliminate all alias components from an image. The 4-step procedure described in the previous section has been performed on a test image containing various patterns. The image is shown in Figure 20a. In Steps 3 and 4 of the procedure the parameters have been reduced until all aliasing errors have been suppressed. The parameters for the best filters have been listed in Table II and the filter magnitude responses are shown in Figure 16. The designed filters have relatively small attenuations, but this has turned out to be sufficient. Higher attenuation only increases the filter order but does not improve the image.

Table II Parameters for anti-aliasing filters for images – "smooth" approach

Parameters	N	ω_p	ω_s
Horizontal filter	14	0.22	0.26
Vertical filter	17	0.16	0.20

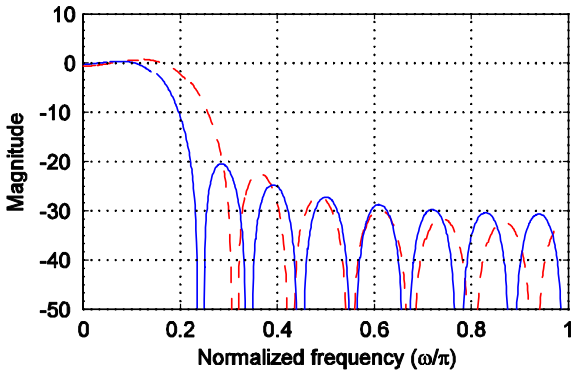


Figure 16. Magnitude response of the horizontal (red dashed line) and vertical (blue solid line) filter for the "smooth" design.

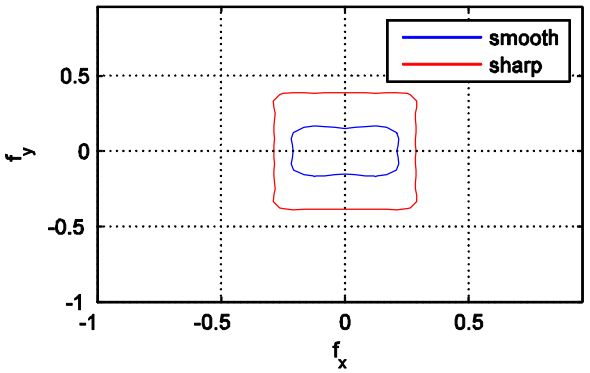


Figure 17. Magnitude response (3dB contour) of the equivalent "smooth" (blue line) and "sharp" (red line) 2D filter.

In order to enable comparison with non-separable filters, in Figure 17 the edges of the passband (3dB point) of an equivalent 2D filter has been shown. The passband surface of this equivalent filter is 0.033.

4.2.2. "Sharp" approach

In the "sharp" approach the goal was to achieve visually good images. Some aliasing was allowed as long as it did not degrade the overall perception of the image. Again, the 4-step procedure described in Section 4.1 has been applied on the image shown in Figure 20a. This time the parameters have been reduced until a 'good' image is obtained. This is highly subjective but as it can be seen in the example section, the selected filters do perform better. The parameters of the best filters are listed in Table III and the filter magnitude responses are shown in Figure 18. The magnitude response (contour) of the equivalent 2D filter has been shown in Figure 17. The passband surface is 0.107. It can be seen that this filter has much wider passband than the non-separable ones as well as the one used for "sharp" approach. After filtering an image with these filters, more details will be preserved but some aliasing will be also visible. However, as it will be illustrated in the example section the gain in image quality is much higher than the visible errors due to aliasing.

It was noticed that the sub-sampling pattern can be approximated by an orthogonal sampling grid rotated clockwise by 12.53 degrees. Based on this fact, it is logically to assume that by rotating the image counter-clockwise by that angle, the sub-sampling pattern would become more regular and as such, more appropriate for filtering. Therefore, in this paper an attempt has been made to do following: First, the image is rotated by 12.53 degrees counter-clockwise. For this purpose, spline-based rotation approach suggested in [8] has been used. Second, the rotated image was filtered with filters designed for the "sharp" approach (parameters are given in Table III). Third, the filtered image has been rotated clockwise by 12.52 degrees. Although, better results were expected, it turned out that filtering rotated or non-rotated images yield to the same visual result. This can be seen in Figure 20d and Figure 20e.

Table III Parameters for anti-aliasing filters for images – "sharp" approach

Parameters	N	ω_p	ω_s
Horizontal filter	22	0.28	0.33
Vertical filter	22	0.38	0.43

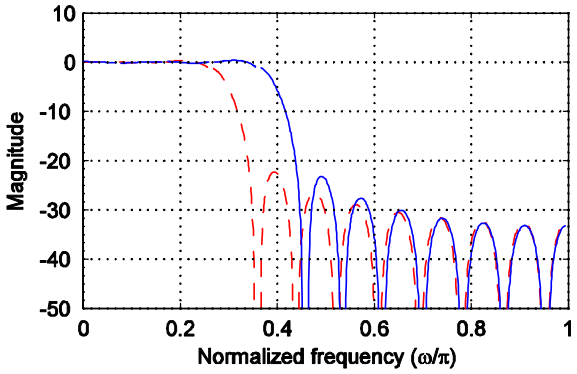


Figure 18. Magnitude response of the horizontal (red dashed line) and vertical (blue solid line) filter for the "sharp" design.

4.3. Anti-aliasing of text

Text can be considered as a special type of an image. Therefore, same filters can be applied as the one discussed in the previous sections. However, when talking about text, readability is the most important property. It has been shown that people often find over-sharpened text to have better readability. For that reason it is better to design a separate filter for processing text as such filter will have even less anti-aliasing properties, than for example the filters designed in the previous section ("sharp" approach). Moreover, it has been experimentally observed that for text only horizontal filtering is enough. This further simplifies the computational complexity of anti-aliasing filtering. The first 3 steps (Step 4 is not needed as only horizontal filters are used) of the procedure described in Section 4.1 have been applied on an image containing only text. The image is shown in Figure 21a. The parameters for the best filter were selected such that the best readability is achieved. These parameters are given in Table IV with the magnitude response of the filter based on these parameters shown in Figure 19. Two points should be emphasized. First, the passband (3 dB edge) occupies 0.285 of the overall band. This is considerably higher than in the case off all previously introduced filters. Second, the passband ripple of this filter is quite high (almost 1 dB). Nevertheless, this is not a problem because we are processing black and white text. Small changes in the brightness are not affecting the readability. Due to the high passband ripple, smaller filter order can be used.

Table IV Parameters for anti-aliasing filters for text

Parameters	N	ω_p	ω_s
Horizontal filter	12	0.30	0.35

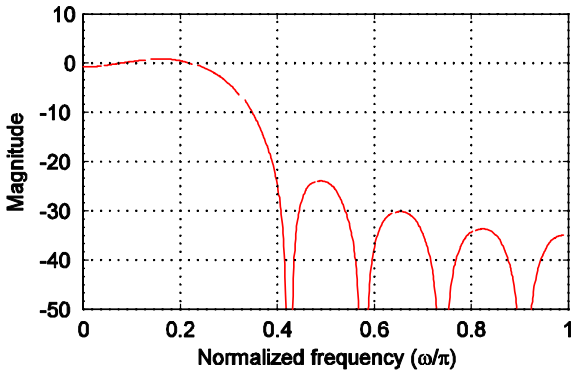


Figure 19. Magnitude response of the horizontal filter for text.

5. RESULTS

Three typical usage scenarios, which result in 2D content being rendered on a 3D display, were identified. Three test images were selected to represent the content in each case. Each image was filtered with different filters, and the results were shown on 23" X3D display and photographed.

The first case is using a 2D graphical user interface to choose or navigate through 3D content. User interfaces contain details with high contrast and straight lines, similar to the "2D graphics" image in Figure 20a. Since the optical filter of 23" X3D is slanted at 12.53 degrees, lines with this orientation are also included in the image. The second case is 2D subtitles being rendered in a 3D movie. This case is represented by the "2D text" test image shown in Figure 21. The "2D text" image contains words with different font sizes and is created using Wordle [11]. The third case is 2D movie being rendered on a 3D display. This is represented by full-colour, natural 2D image. The test image "2D photo" used to test this case is from the Kodak Image Database 0.

In the "2D graphics" tests, there are two groups of filters which produce similar visual results. The first group is filters which remove all aliasing artefacts on expense of over-smoothing the image. The "no crosstalk" non-separable filter (see Figure 20b) and the "smooth" set of separable filters (Figure 20f) fall in this group. The other group is filters which produce sharper image, preserve more details, but leave some aliasing in the image. The "crosstalk aware" non-separable filter (see Figure 20c) and the "sharp" ser of separable filter (Figures 5d and 5e) are in this category. Notably, there is negligible visual difference when applying separable fitters with or without rotation.

In the "2D text" tests, the "text optimized" separable filter produces results with highest readability, as seen in Figure 21d, outperforming the "no crosstalk" and "crosstalk aware" non-separable filters. Finally, in the "2D photo" tests the "sharp" set of separable filters produces best visual results, closely followed by the "crosstalk aware" non-separable filter.

It can be seen from the above figures that visually, the images filtered by the proposed separable filters are considerably better than the one filtered by the 2D filters. Beside this, filtering images with separable filters is also more computationally efficient. In order to illustrate this, in Table V the computational complexity for various filters discussed in this paper are given. As it can be seen, the non-separable filters of size N by N have complexity proportional to N^2, whereas the separable ones of order N have the complexity proportional to N.

Table V Computational complexity for various filters discussed in this paper. C stands for required number of multiplication per pixel.

Filter type		Filter size	C
2D non-separable	one view	48 by 48	2304
	crosstalk	48 by 48	2304

2D separable	“smooth”	15 and 18	43
	“sharp”	23 and 23	46
1D	text	13	13

6. CONCLUSIONS

Different methodologies for design of anti-aliasing filters for multi-view displays were presented. The filters were optimized for 2D content, which is most affected by aliasing artefacts. Two separable and three non-separable anti-aliasing filters were compared for three types of typical 2D content. The effect of the filters was demonstrated on an actual multi-view 3D display. The results show, that specially optimized separable filters can produce similar or better visual results that non-separable ones, while requiring much less computational operations per pixel.

7. ACKNOWLEDGMENTS

This work is partially supported by the European Commission within the ICT Programme of FP7 under Grant 216503 with the acronym MOBILE3DTV.

8. REFERENCES

[1] A. Jain and J. Konrad, “Crosstalk on automultiscopic 3-D displays: Blessing in disguise?,” in *Proc IS&T/SPIE Symposium on Electronic Imaging, Stereoscopic Displays and Applications*, San Jose, CA, Vol. 6490, pp. 649012 (2007).

[2] M. Zwicker, W. Matusik, F. Durand, H. Pfister, "Antialiasing for Automultiscopic 3D displays", In *Proc. of Eurographics Symposium on Rendering*, Cyprus, 2006

[3] S. Pastoor, “3D displays”, in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

[4] A. Boev, A. Gotchev and K. Egiazarian, “Crosstalk measurement methodology for auto-stereoscopic screens”, *in Proc. of 3DTV Con*, Kos, Greece, 2007

[5] C. Van Berkel and J. Clarke, “Characterisation and optimisation of 3D-LCD module design”, *in Proc. SPIE Vol. 2653, Stereoscopic Displays and Virtual Reality Systems IV*, (Fisher, Merritt, Bolas, edts.), p. 179-186, May 1997

[6] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", *in Proc. SPIE Photonics West 2002: Electronic Imaging*, vol. 4660, pp. 212-221, 2002.

[7] X3D-23” Users’ Manual. NewSight GmbH. Firmensitz Carl-Pulfrich-Str. 1 07745 Jena, 2006

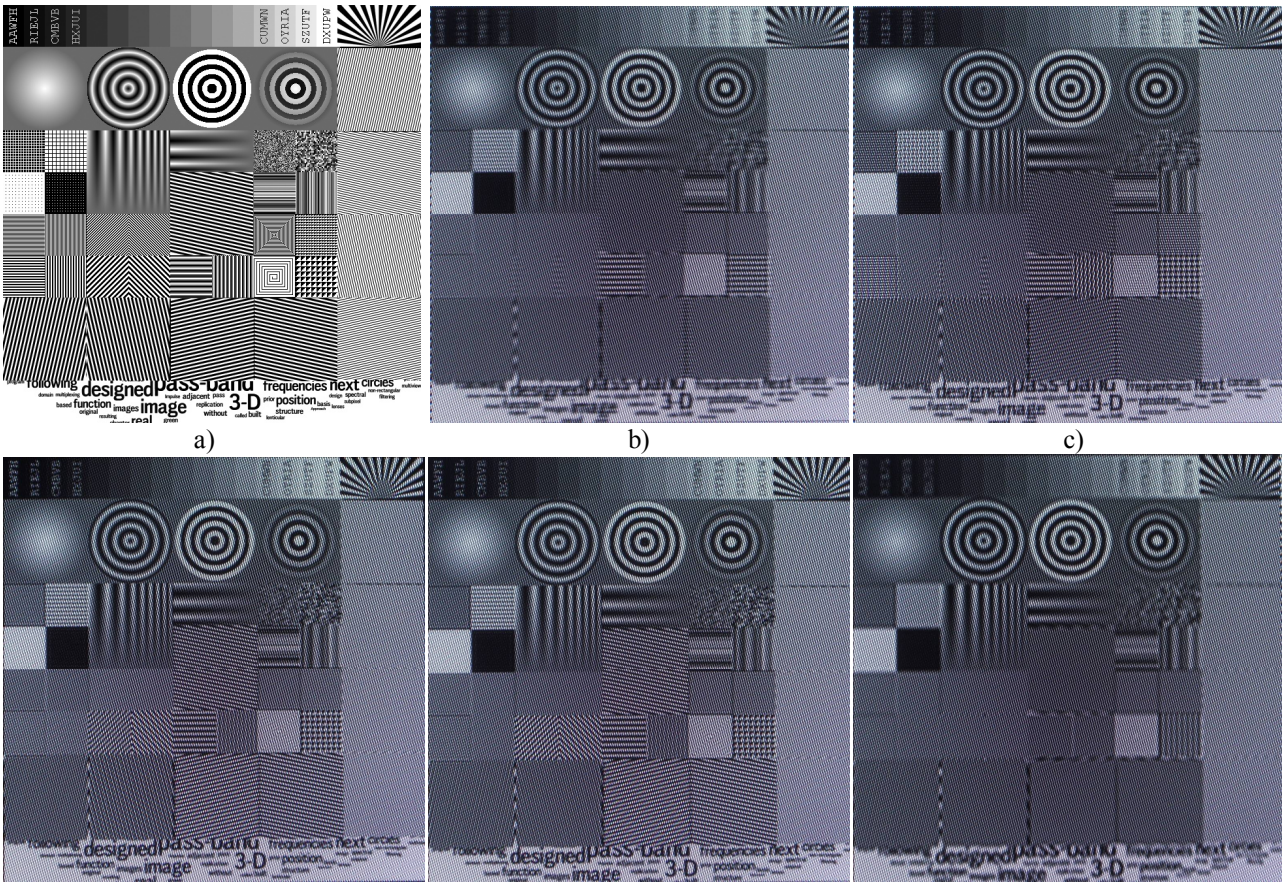
[8] M. Unser, P. Thevenaz, and L. Yaroslavsky, “Convolution-based interpolation for fast, high-quality rotation of images”, *IEEE Trans. Signal Processing*, vol. 4, pp. 1371-1381, 1995.

[9] A. Jain, *Fundamentals of Digital Signal Processing*, Information and System Science Series, Prentice Hall, 1989.

[10] S. K. Mitra, *Digital signal processing: A computer based approach*, New York: McGrow-Hill, 2005

[11] Wordle, online tool for generating “word clouds”, available online at http://www.wordle.net

Kodak image database, available online at ftp://ftp.kodak.com/www/images/pcd



a) b) c)

d) e) f)

Figure 20, “2D graphics” test case: a) original image, b)-f) photographs of the display, as follows – b) filtered with non-separable filter without taking crosstalk into account, c) filtered with non-separable filter taking crosstalk into account, d) filtered with “sharp” set of separable filters after rotation, e) filtered with “sharp” set of separable filters without rotation, and f) filtered with “smooth” set of separable filters without rotation

a) b)

c) d)

Figure 21, “2D text” test case: a) original image, , b)-d) photographs of the display as follows – b) filtered with non-separable filter without taking crosstalk into account, c) filtered with non-separable filter taking crosstalk into account, and d) filtered with set of separable filters optimized for text, without rotation

a) b) c)

d) e) f)

Figure 22, “2D photo” test case on multi-view display: a) original images, b)-f) photographs of the display, as follows – b) without filter, exhibiting crosstalk, c) filtered with non-separable filter without taking crosstalk into account, d) filtered with non-separable filter taking crosstalk into account, e) filtered with “sharp” set of separable filters without rotation, and f) filtered with “smooth” set of separable filters without rotation

[P08] A. Boev, D. Hollosi, Atanas Gotchev and Karen Egiazarian, "Classification and simulation of stereoscopic artifacts in mobile 3DTV content", S*tereoscopic Displays and Applications XX, Proc. SPIE 7237*, 72371F (2009), DOI:10.1117/12.807185

Correction: Word in Fig. 3, second ring, top. Correct: “aberration”. Labels on Fig. 8, bottom row. Correct: “d) e) f)”. Section 5, paragraph 3, line 5, word 6. Correct: “stream”.

This page left blank

Classification and simulation of stereoscopic artifacts in mobile 3DTV content

Atanas Boev, Danilo Hollosi, Atanas Gotchev, Karen Egiazarian
Institute of Signal Processing, Tampere University of Technology, P.O.Box 553, 33101 Tampere, Finland

ABSTRACT

We identify, categorize and simulate artifacts which might occur during delivery stereoscopic video to mobile devices. We consider the stages of 3D video delivery dataflow: content creation, conversion to the desired format (multiview or source-plus-depth), coding/decoding, transmission, and visualization on 3D display. Human 3D vision works by assessing various depth cues – accommodation, binocular depth cues, pictorial cues and motion parallax. As a consequence any artifact which modifies these cues impairs the quality of a 3D scene.

The perceptibility of each artifact can be estimated through subjective tests. The material for such tests needs to contain various artifacts with different amounts of impairment. We present a system for simulation of these artifacts. The artifacts are organized in groups with similar origins, and each group is simulated by a block in a simulation channel. The channel introduces the following groups of artifacts: sensor limitations, geometric distortions caused by camera optics, spatial and temporal misalignments between video channels, spatial and temporal artifacts caused by coding, transmission losses, and visualization artifacts. For the case of source-plus-depth representation, artifacts caused by format conversion are added as well.

Keywords: mobile 3DTV, mobile 3D video, stereoscopic artifacts, stereoscopic video quality, portable 3D displays

1. INTRODUCTION

Recently, most of the building blocks of an end-to-end mobile 3DTV system have reached maturity status. An ISO/MPEG multiview encoding standard developed as an amendment to H.264 AVC is being standardized[1, 2]. Various algorithms have been developed for the efficient transmission of video streams over wireless networks[1, 3]. There are 3D displays optimized for a mobile use[4, 5, 6.] While the core technologies have been developing, there is still much to be done to optimize the system to deliver the best possible visual output[7, 8.] Having a perceptually acceptable and high-quality 3D scene on a small display is a challenging task.

Estimation of the quality is the key factor in design and optimization of any visual content. All quality metrics aim at close approximation of the quality as perceived by the user. An ideal quality metric should have the following properties: a) perceptual – being related to the way human visual system (HVS) operates, b) objective – providing a numerical representation of the quality as perceived by the user, and c) reliable – being able to predict the perceptual quality for wide variety of content, as perceived by a large amount of users. Such metric is especially needed for stereoscopic 3D video, because stereoscopic artifacts would produce not only visually unpleasant results, but are also known to cause eye-strain general discomfort[9]. The previous works on quality of stereo images[10, 11] do not attempt to quantify the typical distortions that could occur in stereoscopic video sequence.

The first step towards objective quality estimation metric is to identify the artifacts, which could occur in various usage scenarios. Then, subjective tests should be performed, in which human observers would grade the perceptual quality of a variety of content. In this work, we attempt to identify the artifacts, which could occur in a mobile 3DTV system. We present a system which allows a set of stereoscopic artifacts to be introduced to a given 3D video, thus ensuring repeatability of subjective quality experiments. In the next section, we discuss the "layered" nature of the human 3D vision. In chapter 3 we introduce a concept for broadcasting stereo-video over DVB-H channel, and describe which stages of such system can introduce artifacts. In section 4, we compare the delivery stages to the "layers" of 3D vision to build a classification of stereoscopic artifacts. In section 5 we present a framework for simulation of mobile 3DTV artifacts. Finally, section 6 describes the mobile 3DTV artifacts simulated by our framework.

Stereoscopic Displays and Applications XX, edited by Andrew J. Woods, Nicolas S. Holliman, John O. Merritt, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 7237, 72371F

2. PERCEPTION OF DEPTH

The human visual system is a set of separate subsystems, which operate together as a single process. It is known that spatial, color and motion information is transmitted to the brain using largely independent neural paths[12]. Vision in 3D, in turn, also consists of different "layers" which provide separate information about depth of the observer scene[12, 13]. This is true both for perception and cognition – on perceptual level there are separate visual mechanisms and neural paths, and on cognitive level there are separate families of depth cues, with varying importance from observer to observer[12, 14]. The depth cues used in different layers in human vision are illustrated in Figure 1 and are as follows:

- Accommodation – This is the ability of the eye to change the optical power of its lens in order to focus on objects at various distances. Accommodation is the primary depth cue for very short distances, where an object is hardly visible with two eyes. With the distance, the importance of this depth cue quickly decreases. However, the information from other depth-assessing systems is unconsciously used to correct the refraction power, to ensure clear image of the object being tracked. As a result, a discrepancy between accommodation and binocular depth cues leads to so called *accommodation-convergence rivalry*, which is a major limiting factor for stereoscopic displays.
- Binocular depth cues – These are a consequence of both eyes observing the scene at slightly different angles. The mechanism of binocular depth estimation has two parts – *vergence* and *stereopsis*. Vergence is the process, in which both eyes take a position which minimizes the difference of the visual information projected in both retinae. The angle between the eyes is used as a depth cue. With the eyes converged on a point, stereopsis is the process which uses the residual disparity of the surrounding area for depth estimation relative to the point of convergence. Binocular depth cues are the ones most often associated with "3D cinema". However, binocular vision is quite vulnerable to artifacts – lots of factors can lead to an "unnatural" stereo-pair being presented to the eyes. As HVS is not prepared to handle such information, binocular artifacts can lead to nausea and *simulator sickness* or *cyber sickness* [9]. It is worth saying, that around 5% of all people are "stereoscopically latent" and have difficulties assessing binocular depth cues[11, 13]. Such people have a perfect depth perception, but rely mostly on depth cues coming from other visual "layers".
- Pictorial cues – for longer distances, binocular depth cues become less important, and HVS relies on pictorial cues for depth assessment. These are depth cues that can be perceived with a single eye – shadows, perspective lines, texture scaling. But even for medium distances, stereoscopically good image can be "ruined" if missing subtle pictorial details, and the scene exhibits *puppet theatre* or *cardboard effect* artifacts
- Motion parallax – this is the process in which the changing parallax of a moving object is used for estimating its depth and 3D shape. The same mechanism is used by insects, and is commonly known as "insect navigation"[15]. Artifacts in the temporal domain (e.g. motion blur, display persistence) will affect the motion parallax depth cues.

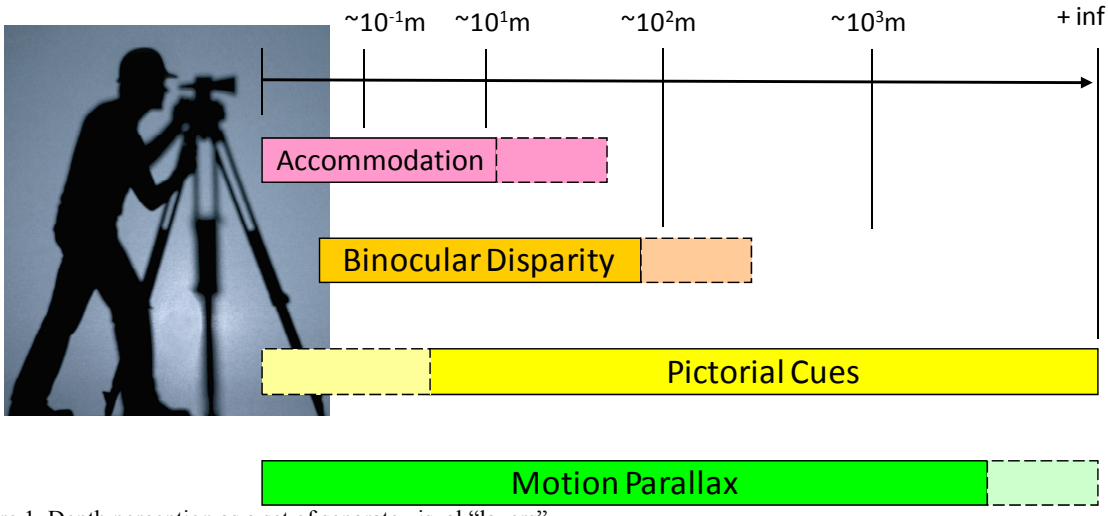


Figure 1, Depth perception as a set of separate visual "layers"

Experiments with so-called "random dot stereograms" show that binocular and monocular depth cues are independently perceived[16]. Furthermore, the first binocular cells (cells that react to a stimulus presented to either of the eyes) appear at a late stage of the visual pathways – the V1 area of brain cortex. At this stage, only the information extracted separately for each eye, is available to the brain for deduction of image disparity[12]. This observation has led us to the assumption that "2D" (monoscopic) and "3D" (stereoscopic) artifacts would be independently perceived[17]. The planar "2D" artifacts, such as noise, ringing, etc, are thoroughly studied in the literature[18]. Here, we focus on artifacts which affect stereoscopic perception. However, due to the "layered" structure of HVS, binocular artifacts might be inherited from other visual "layers" – for example, *blockiness* is a "purely" monoscopic artifact, which still can destroy or modify an important binocular depth cue.

3. ARTIFACTS IN MOBILE 3DTV SYSTEM

The dictionary describes *artifact* as "something characteristic of or resulting from a human institution or activity"[19]. Non-natural processes, as is the case of transmitting a 3D scene representation over a communication channel, are source of artifacts.

One case of such transmission is a mobile 3DTV system, where a 3D (usually stereoscopic) video stream is broadcasted over the air and received on a portable device. In such system, stereoscopic video content is captured, encoded, encapsulated and then broadcast over a DVB-H channel, and is received, decoded and played by a DVB-H enabled portable device with autostereoscopic display. The data flow from creation to observation is shown in Figure 2.

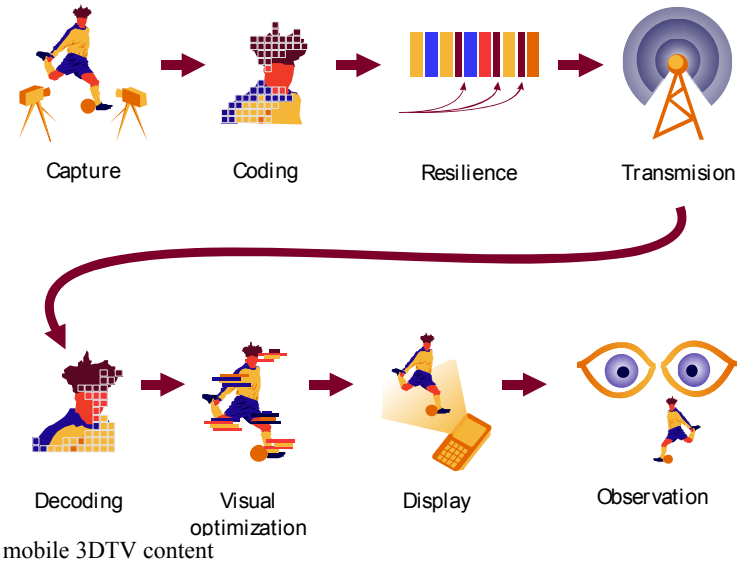


Figure 2, Data flow of mobile 3DTV content

The stages of the dataflow can create various artifacts as follows:

- Creation/capture –three common approaches for 3D video capture. First, such video can be captured by two or more synchronized cameras in a multi-camera setting. Second, such content can be created from 2D video applying video processing methods. Third, video output can be augmented by depth information captured by another sensor. All these approaches have their own advantages and disadvantages, and are sources of specific artifacts. Special care should be taken when positioning cameras or when selecting rendering parameters. Unnatural correspondences between the images in a stereo-pair (i.e. vertical disparity) are source of many types of artifacts[11]. As perfectly parallel camera setup is practically impossible, rectification is an unavoidable pre-processing stage.

- Representation format – Although there are many different formats for encoding 3D video, three main groups have evolved: *multiview video*, where two or more video streams show the same scene from different viewpoints; *video-plus-depth*, where each pixel is augmented with information of its distance from the camera; and *dynamic 3D meshes*, where 3D video is represented by dynamic 3D surface geometry [20]. Video-plus-depth format is suitable for multiview displays, as it can be used regardless of the number of views a particular screen provides[20]. On the downside, video-plus-depth rendering requires inpainting of occluded areas, which causes *disocclusion artifacts*. This problem has been addressed by using layered depth images (LDI), or multiview video-plus-depth encoding[21].

If the representation format is different from the one the scene was originally captured, format conversion is another source of artifacts. Some artifacts which are common in one format and not possible in another – for example in video-plus-depth disocclusion artifacts are common, while vertical parallax does not occur.

- Coding – there are various coding schemes, which utilize temporal, spatial or inter-channel similarities of a 3D video[20]. Two approaches are most popular for stereo-video – multi-view coding, standardized as an amendment to H.264/AVC [1, 2]; and 2D video with separate depth channel, which can be compressed using H.264/AVC and stored in MPEG container [22, 23]. Special care should be taken when algorithms originally designed for single video channel are used for stereoscopic video, as important binocular depth-cues may be lost.
- Transmission – in the case of digital wireless transmission a common problem is burst packet losses[24]. Resilience and error concealment algorithms attempt to mitigate the impact on the video, but if not designed for stereo-video, such algorithms might introduce additional artifacts on their own.
- Visualization – there are many approaches for 3D scene visualization, offering different degree scene approximation [25]. Each family of 3D displays has its own characteristic artifacts, and the artifacts are often scene dependant [7, 11].

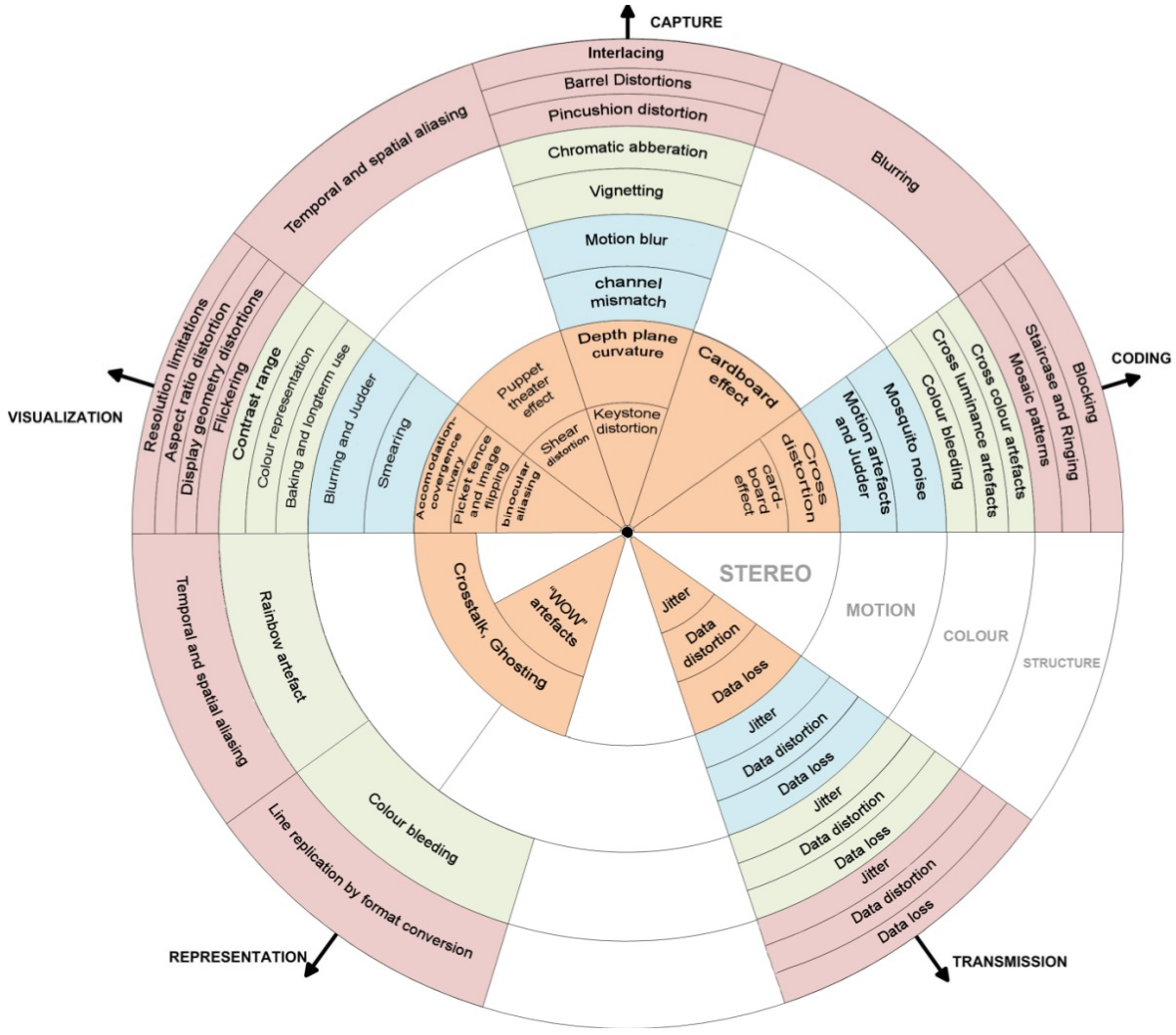


Figure 3, Artifacts, caused by various stages of content delivery and affecting various "layers" of human depth perception.

As a result, stereoscopic artifacts might be created during various stages in the mobile 3DTV content delivery, and might affect different "layers" of human 3D vision, as shown in Figure 3.

4. ARTIFACT CLASSIFICATION

In 3D video, many causes might lead to unnatural scene representations. For building taxonomy of stereoscopic artifacts, we use a top-down approach: first we identify content delivery stages, which might create artifacts, and then we speculate if and how these artifacts will affect various stages of human perception of depth. Our classification is presented in Table 1. The columns represent the causes for artifacts, coming from different content delivery stages – capture, representation, coding, transmission and visualization. The rows are groups of artifacts as they are interpreted by the "layers" of human vision – structure, color, motion and binocular. These layers roughly represent the visual pathways as they appeared during the successive stages of evolution. By *structure* we denote the spatial (and color-less) vision. It is assumed that during the evolution human vision adapted for assessing the "structure" (contours and texture) of images[26], and some artifacts manifest themselves as affecting image structure. *Color* and *motion* rows represent the color and motion vision, accordingly. As we noted before, all artifacts in the table affect the binocular depth perception. However, the row designated with *binocular* contains artifacts which have meaning only when perceived as a stereo-pair. In other words, these are artifacts that cannot be perceived with a single eye (e.g. vertical disparity).

Table 1 – Classification of stereoscopic artifacts

	Capture	Representation/ Conversion	Coding	Transmission / Error Resilience	Visualization
STRUCTURE	- blur by defoucusing - barrel distortions - pincushion distortions - interlacing - temporal and spectral aliasing - downsampling - noise introduction	- temporal and spatial aliasing - line replication	- blocking artefacts - mosaic patterns - staircase effect - ringing	- data loss - data distortion - jitter	- flickering - resolution limitations - aspect ratio distortions - display geometry distortions - spatial aliasing by subsampling on non-rectangular grid)
COLOR	- Chromatic aberration - Vignetting- decressing intensity	- temporal and spatial aliasing	- cross- colour artefacts - colour bleeding	- color bleeding	- contrast range - colour representation - baking and longterm use - viewing angle dependant colour representation - rainbow artefact
MOTION	- motion blur - temporal mismatch		- motion compensation artefacts - mosquito noise - Judder	- loss/distortion in motion - jitter	- smearing - bluring and judder
BINOCULAR	- depth plane curvature - keystone-distortion - cardboard effect	- ghosting by disocclusion - Perspective-binocular rivalry ("WOW"- artefacts)	- cross distortions - cardboard effect - depth "bleeding"/depth "ringing"	- data loss, one channel - data loss, propagating	- shear distortion - ghosting by crosstalk - angle dependant binocular aliasing - accomodation convergence rivalry - lattice artefacts - puppet theater effect - picket fence effect - Image flipping (Pseudoscopic Image)

The process of artifact mapping is not always straightforward – sometime one stage in the dataflow might cause several types of artifacts, while sometimes artifacts created in different stages are perceived in a similar way (e.g. ghosting). Following that, the diagram from Figure 3 cannot be easily translated to a flat table. Some artifacts are listed repeatedly, while some artifacts groups span across multiple cells. Furthermore, some combinations of rows (causes for artifacts) and columns (artifact manifestations) are omitted as unrelated to the usage scenario of mobile 3DTV.

5. ARTIFACT SIMULATION FRAMEFORK

Not all of the stereoscopic artifacts are likely to affect a mobile 3DTV system. Some of them are not applicable to mobile device due to the technology used (e.g. LCD display, DVB-H transmission). Others cannot be mitigated through the means of signal processing, and are usually addressed by content providers and/or display manufacturers.

We suggest an artifact simulation channel, which is able to introduce an arbitrary combination of artifacts to a video, with controlled amount of impairment for each artifact. Artifacts are organized in groups, which follow the flow of a mobile 3D video over a DVB-H channel[28]. Each group of artifacts corresponds to a specific block of our simulation channel as shown in Figure 4. An arbitrary combination of artifacts can be introduced by this channel, but they are always introduced in a certain order – i.e. capture artifacts will always be added before transmission ones.

The first block simulates artifacts caused by sensor limitations. Then, the degraded scene observation is sent to a block which simulates geometric distortions as the ones caused by the camera optics. The next two blocks add global spatial and temporal differences between the video channels, simulating artifacts caused by multi-camera topology and temporal misalignment. The next two blocks simulate spatial and temporal artifacts caused by coding. Then, transmission losses are simulated in the encoded scream. For the case of 3D video represented as color video channel augmented with per-pixel depth information (also known as source-plus-depth video), format conversion artifacts are added. Finally, visualization artifacts are added, independent of the position of the observer, or alternatively, for a given observation position.

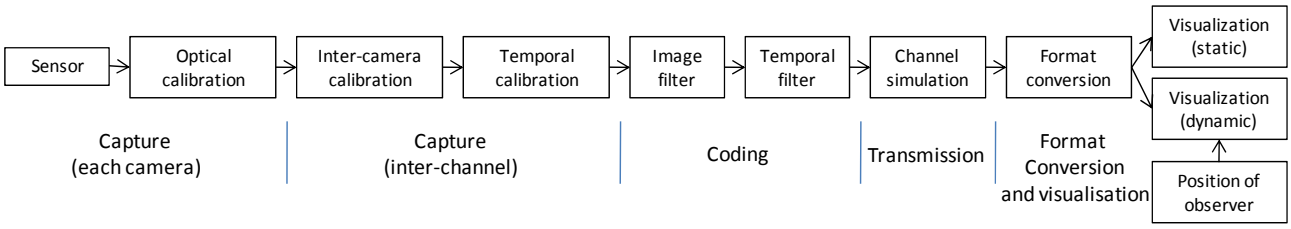


Figure 4, Artifact simulation channel

Following this concept, we have developed a framework for simulation of mobile 3DTV artifacts. The framework is thoroughly described in[29]. It is organized as a collection of Matlab functions, each one responsible for introducing a specific artifact. Additionally, there is a program module, which executes the simulation functions as prescribed by a configuration file. The configuration is stored in a text file, which describes the input and output video streams, the set of artifacts to be introduced, and the parameters for each artifact. One configuration file can specify a set of artifact parameters, which to be applied over several input video files in "batch mode".

The framework operates on stereo-video streams (where left and right channel are provided as separate video files) or source-plus-depth video streams (where video and depth channel are provided as two separate video files). Video is decoded into a set of frames, each frame is processed and the result is encoded in a video stream again. The blocks of the framework are shown in Figure 5, and are as follows:

- GUI – provides two alternative ways to prepare a configuration file – using Microsoft Excel sheet or using Matlab GUI. The Excel-based GUI uses VBA-scripting. Alternatively, a configuration file can be prepared using a text editor [29].
- Session manager – opens and parses a set of configuration files.
- General logic – imports video streams or collections of frames; processes them as described in the configuration file; exports video stream or a set of frames.
- Low level processing – introduces artifacts to a given video by processing each frame separately. While applying artifact to one frame, information from other frames or video channels might be used.
- Database of artifact simulation functions – a set of functions, each one responsible for introducing a specific artifact. Most of the functions are implemented on Matlab. Three functions – "2D+Z to multiview conversion", "Multiview to 2D+Z conversion" and "DVB-H packet loss simulation" are implemented as Windows executables, and are called as external functions by the framework.

The next section describes the full list of artifacts, which can be introduced by our framework.

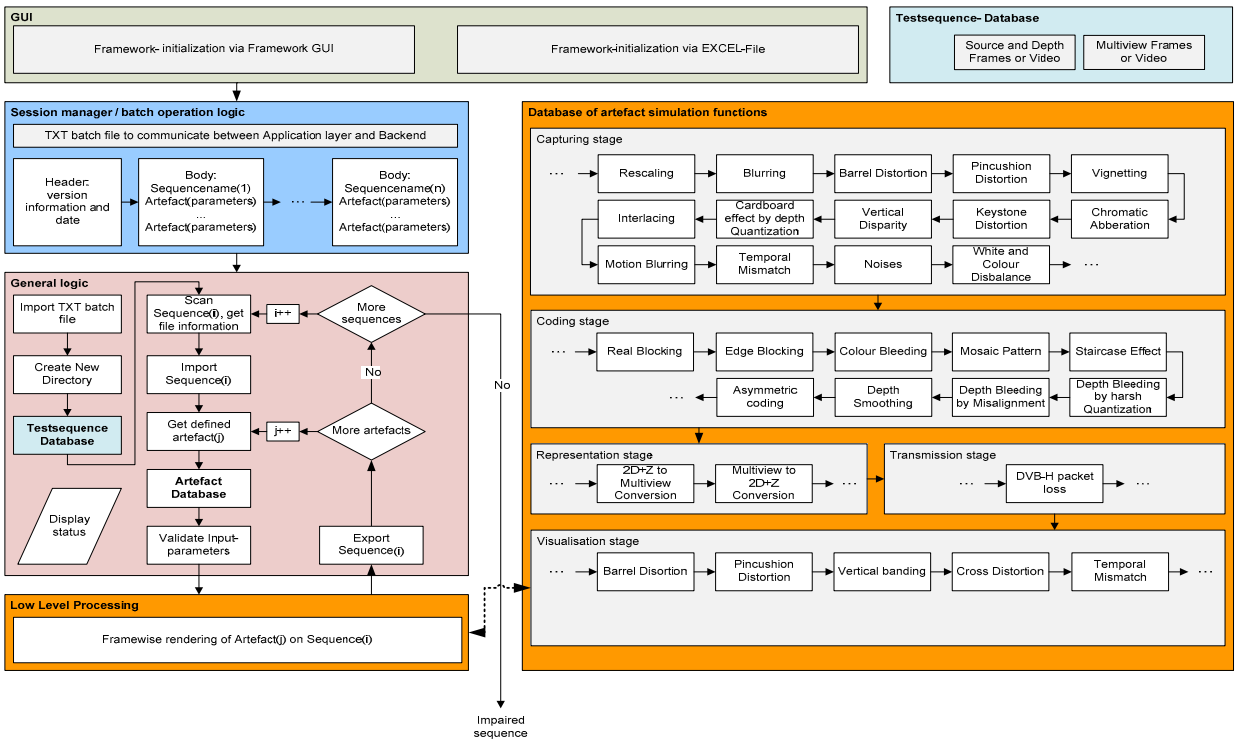


Figure 5, Block diagram of the artifact simulation framework

6. MOBILE 3DTV ARTEFACTS SELECTED FOR SIMULATION

6.1 Capture artifacts

The capturing process for mobile 3DTV video is similar to the one for a 3DTV system targeting large displays. One thing which separates video broadcast system from video conferencing one is that capture for the former is done off-line and non-real-time, and significant processing power might be spent for producing the best output possible.

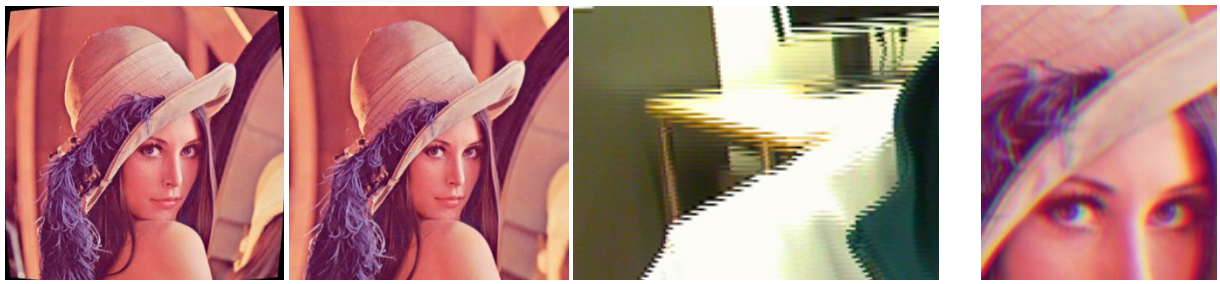
We have chosen for simulation the following list of common stereo video capture artifacts:

- *Size and resolution changes* – the problem of choosing the proper resolution for capturing of 3D contents is not necessarily a simple one. Two problems might arise from content resizing – aliasing and wrong disparity range. The perceptual impact of aliasing on stereoscopic video is yet to be studied – if is it going to be masked by binocular suppression, or is going to destroy important texture-based binocular ques. Additionally, changing the size of a multiview 3D video changes the inter-channel relations as well, which might result in a disparity either too small or too large for proper 3D effect. Our framework allows rescaling of video content using various interpolation methods.
- *Blur* might be caused by low-quality optics or wrong focal setting. In a 2D movie, in most cases small amount of blur is permissible. In a binocular setup, predicting how blur will affect quality is more complex task. Depending on the case, blur in one channel might go unnoticed, or in rare cases even improve the perceived quality.
- *Motion blur* – this is usually caused by capturing in low light conditions. The temporal masking and perception of motion blur in stereo video is yet to be studied.
- *Barrel/Pincushion distortion* is a geometrical distortion, which affects each camera separately. In multi-camera it could cause serious artifacts in stereoscopic image, and induce eye-strain. This is corrected by rectification. These artifacts are simulated by applying identical geometric transformation separately to each channel.
- *Keystone distortion* affects the geometric relation between two channels. The result is a trapezoidal shape in opposite direction in left and right camera inputs. It is mainly caused by camera optics and selected multi-camera

topology. The presence of keystone distortion can induce eye-strain or fully break the 3D effect of a stereo video. It also will greatly diminish the precision of dense depth estimation algorithms. Image rectification compensates this effect. Keystone distortion is introduced by simulation of converging camera setup – namely applying projective distortion to each channel, with opposite projection directions.

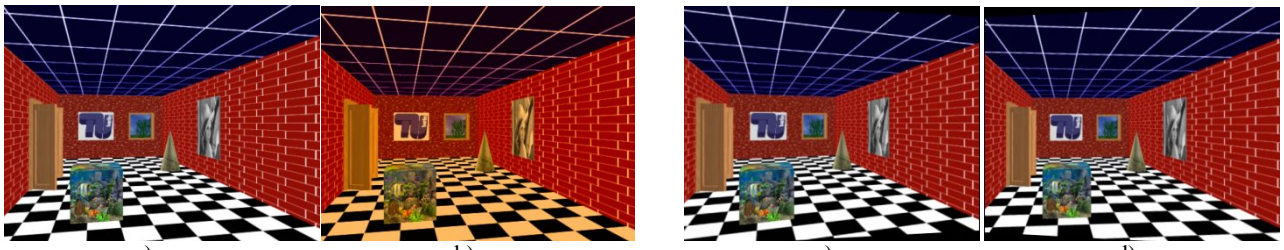
- *Temporal mismatch* occurs when a 3D scene is shot with multiple cameras, which are not shutter-synchronized. As a result, the frames in both channels are not shot simultaneously, but with slightly shifted in time. While precise time synchronization is of crucial importance for dense depth estimation algorithms, the human visual system can tolerate some amount of time mismatch, without diminishing the perceptual quality. We simulate temporal mismatch by adding temporal offset to one of the channels.
- *Color mismatch* – Some factors (i.e. bright objects with large disparity between cameras) can cause mismatch in the colors in the images of a scene captured by different cameras. It is most commonly caused by white balance done separately in each camera. Color mismatch is introduced by mimicking automatic white balancing algorithms, with selectable illumination parameters.
- *Interlacing* – Interlaced video is created by scanning the odd and the even lines of an image sensor separately. Interlaced video exhibits specific "jagged-border" artifacts as seen in Figure 6. In 2D video, interlacing overlaps consecutive frames in time. As some stereo-video encoding methods involve using of odd and even fields, interlacing might also interleave simultaneous frames from different channels.
- *Cardboard effect* refers to unnatural flattening of objects in stereoscopic images, as if they were cardboard cut-outs[12]. It is believed that the main reason is the field of view of a stereoscopic display being different from the field of view of the scene, thus creating inappropriate depth scaling[14]. In our framework we simulate cardboard effect only on video streams represented in source-plus-depth format. However, our framework could be extended to simulate cardboard effect in other video formats.

Additionally, we simulate less common capture artifacts such as *noise*, *vignetting* and *chromatic aberration*. Proper simulation of *camera noise* is a very demanding task, but usually it is dealt separately inside each camera. We included noise simulation for the ability to prepare subjective test material where asymmetric amount of noise is present in each channel.



a) b) c) d)

Figure 6, Example of monoscopic 3D artifacts introduced during capture: a)barrel distortion, b)pincushion distortion, c)interlacing and d)color aberration



a) b) c) d)

Figure 7, Example of stereoscopic 3D artifacts introduced during capture: a),b) stereoscopic pair exhibition color mismatch, and c), d) stereoscopic pair with added keystone distortion

6.2 Coding artifacts

While the visibility of coding artifacts is quite well studied for 2D case, the impact on the 3D vision is yet to be determined.

- Transform-caused artifacts come from the transforms and quantization used for compressing the video stream. *Blocking, mosaic patterns, staircase effect, ringing,* and *color bleeding* artifacts are in this group. All of them are well visible, and as overlay structural changes on the image, they might destroy depth cues and even create misleading ones. *Depth bleeding* and *depth ringing* are artifacts specific for the coding the depth map of a scene, and as such, they exist only in source-plus-depth-based 3D video representations. Notably, such artifacts can be mitigated by using structural information of the 2D scene.
- Temporal coding artifacts appear as a result of transform/quantization over time. Temporal inconsistency such as *mosquito* noise is the most common artifact in this group. Artifacts caused by imprecise *motion prediction* are also possible. This group of artifacts can appear both in multi-view and in source-plue-depth 3D video.

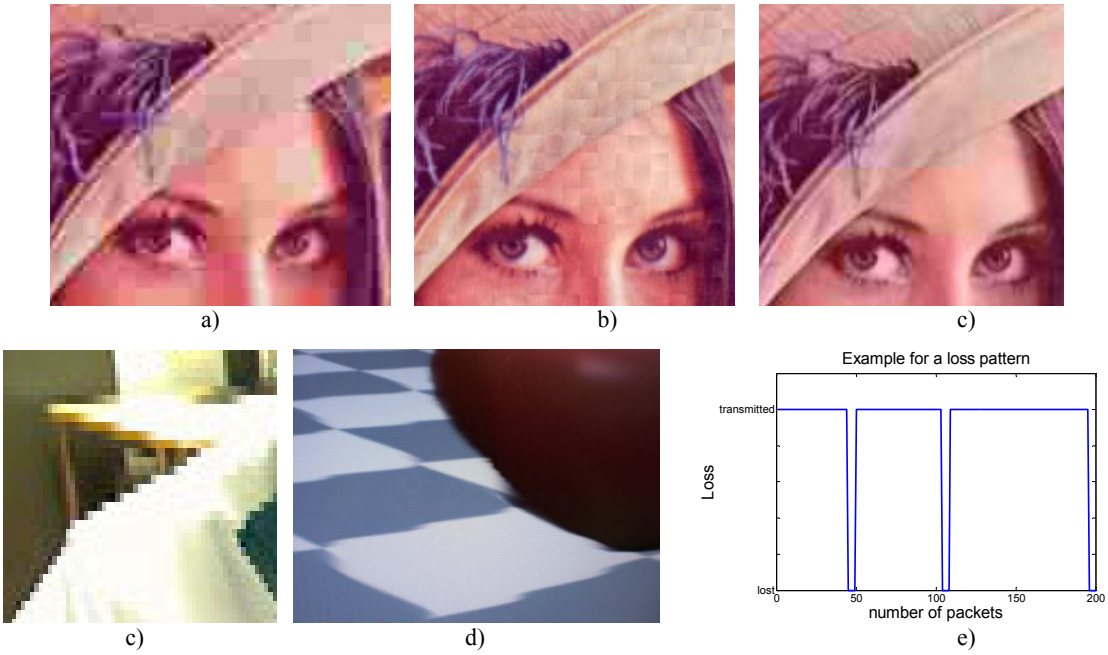


Figure 8, a)-d) Example of coding artifacts: a) blocking by harsh quantization, b) blocking by edge discontinuities, c) clolr bleeding, d)staircase effect, depth bleeding; e) example error pattern of DVB-H channel losses.

Our framework simulates the following coding artifacts:

- *Blocking by harsh quantization* is among the most widely studied distortions in video coding. The most common source of the artifact is block-based DCT compression, which involves quantization of the results. This process creates a number of image impairments, the most noticeable of which are discontinues at the boundaries of the encoded blocks. In our framework, we simulate blocking by harsh quantization by utilizing the DCT block-based compression used in JPEG. The results are seen in Figure 8a. Additionally, some authors propose that blocking might be considered as several, visually separate artifacts – *block-edge discontinuities*, *color bleeding*, *blur* and *staircase artifacts*[18]. Our artifact also provides means to simulate these artifacts separately, if needed.
- *Block-edge discontinuities* – block-based coding tries to exploit the spatial correlation between pixels in a picture, but does not take into account the possible correlation beyond the block borders. One important property of such distortion is that the mean intensity of the block remains the same as before. In our framework, we provide an option that block-edge discontinuities are simulated separately from block-based DCT artifacts. We simulate block-edge discontinuities by introducing luminance distortions inside of a block while keeping the mean luminance constant, as seen in Figure 8b.

- *Color bleeding* is an artifact caused by harsh quantization of high frequency chrominance coefficients. Since chrominance is typically sub-sampled, bleeding can occur beyond the range of a block. Color bleeding is simulated by applying different levels of quantization to chrominance and luminance channels, as illustrated in Figure 8c.
- *Staircase effect* affects diagonal edges of a picture. The quantization of DCT coefficients results in diagonal lines which are almost horizontal or almost vertical, to be represented by a series of blocks. We approximate the effect of staircase artifact by selective pixel doubling in horizontal or vertical direction, which produces staircase edges as the ones seen in Figure 8c.
- *Cross-distortion* is an artifact caused by asymmetrical stereo-video coding. The asymmetry might be both in spatial (one channel with lower resolution) or in temporal (one channel having lower frame-rate) domains. The effect of spatial or temporal sub-sampling of one channel is not yet thoroughly studied. Asymmetrical coding is applied for multi-view video only.

Additionally, our framework simulates less common coding artifacts which affect videos in image plus depth format – *depth bleeding* and *depth smoothing*. Depth bleeding is caused by a process similar to the one, which causes color bleeding – with the difference that it degrades the depth channel instead of the chrominance. Depth smoothing could be caused by asymmetric compression or rescaling of the depth channel. In some cases, depth smoothing might improve the quality of an image plus depth video, as it will hide some disocclusion artifacts.

6.3 Conversion artifacts

Format conversion artifacts occur during the conversion for a source-plus-depth representation used for broadcast to a multiview one as needed by the display. Most common here are *disocclusion artifacts*, which are more pronounced when rendering observations at angles much different from the central observation point, and less pronounced when layered depth images are used[21]. *Perspective-stereopsis rivalry* occurs if the conversion over-exaggerates the depth levels in the depth map. *Temporal inconsistency* of the depth estimation creates artifacts similar to mosquito and depth ringing. It is quite difficult to simulate conversion artifacts separately from the actual process of conversion. Our framework allows various types of conversion algorithms and quality settings to be used for introducing of conversion artifacts.

6.4 Transmission artifacts

The presence of artifacts generated in the transmission stage depends very much on the coding algorithms used and how the decoder copes with the channel errors. In DVB-H transmission most common are burst errors[27], which results in packet losses distributed in tight groups. In MPEG-4 based encoders packet losses might result in propagating or non-propagating errors, depending on where the error occurs in respect to the I-frames, and the ratio between I- and P-frames. We simulate transmission errors by obtaining error patterns of the DVB-H channel and use them for simulation of channel losses as it is done in[27]. Example DVB-H error pattern is shown in Figure 8e.

6.5 Visualization artifacts

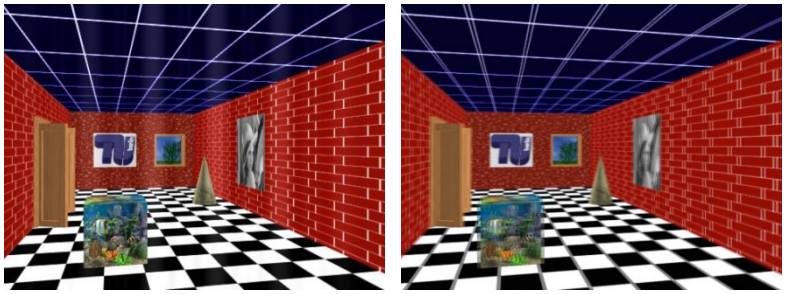
Artifacts in visualization of mobile 3DTV are caused by limitations of the display technology used. We expect a mobile 3DTV system to use 2-view, autostereoscopic display. Such displays use spatial multiplexing of the channels, and the visibility of all artifacts depends on the position of the observer.

Some visualization artifacts are perceived while changing the position in respect to the display. Such artifacts are *angle dependant color representation*, *pseudostereoscopy*, *picket fence effect*, or the unnatural image parallax causing *shear distortion*. Others appear only for some observation angles, as *image flipping*. The artifacts in this group are difficult to simulate, but easy to mitigate for a given position of the observer.

In our framework, we simulate only artifacts, visible by a static observer:

- *Vertical banding* can be regarded as the "static" version of picket fence effect. It is very common for displays with parallax barrier, and manifests itself as changes of the intensity across the display – as if dark vertical bands are superimposed on the image. Even though it depends on the viewing angle, it is visible from most of the viewing angle/observation distance combinations, except for a few observation "sweet spots". An example of simulated vertical banding can be seen in Figure 9a.

- *Temporal mismatch* is a temporal misalignment between the video channels. While during capture such misalignment is usually very small, but depending on the decoder, temporal misalignment can increase to several seconds. Typical causes are reception problems and rudimentary error concealment.
- *Resolution change*– it is possible that a stereo-video stream needs to be rescaled on the receiving device. Rescaling can create aliasing and improper disparity, similarly to the effects during capture. Additionally, rescaling during visualization might affect (exaggerate or suppress) other artifacts.
- *Cross-talk* – display imperfections can cause cross-talk and other forms of inter-channel distortion. Stereo- and multiview displays using parallax barrier are particularly vulnerable to crosstalk. Crosstalk is simulated by scaling the intensity of a frame in one channel and superimposing it over a frame of the other channel, as seen in Figure 9b.



a) b)

Figure 9, Example of 3D artifacts introduced in the visualization stage: a) banding artifacts, and b) crosstalk.

7. CONCLUSION

We identified which 3D artifacts can occur in a mobile 3DTV system, featuring H.264 AVC type of encoding, DVB-H transmission channel and portable autostereoscopic display. We discussed how different stages of mobile 3DTV content delivery could affect the subsystems of human 3D vision. We proposed artifact simulation channel which follows the natural flow of a mobile 3D video over a DVB-H channel.

We presented an artifact simulation framework that allows an arbitrary combination of artifacts to be introduced to 3D video. Such framework can be used to perform subjective experiments, in which the perceptual quality of various mobile 3DTV artifacts can be estimated.

8. ACKNOWLEDGEMENT

This work is supported by the European Commission within the ICT Programme of FP7 under Grant 216503 with the acronym MOBILE3DTV.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11, “Study Text of ISO/IEC 14496-10:2008/FPDAM 1 Multiview Video Coding”, Doc. N9760, Archamps, France, May 2008.

[2] ISO/IEC JTC1/SC29/WG11, “Joint Multiview Video Model 8”, Doc. N9762, Archamps, France, May 2008.

[3] Z. Tan and A. Zakhor, “Error control for video multicast using hierarchical FEC," in *Proc. of the Int. Conf. on Image Processing*, Kobe, Japan, October 1999, vol. 1, pp. 401-405.

[4] G. J. Woodgate, J. Harrold, “Autostereoscopic display technology for mobile 3DTV applications”, in *Proc. SPIE* Vol.6490A-19, 2007

[5] Sharp Laboratories of Europe, website, http://www.sle.sharp.co.uk/research/optical_imaging/3d_research.php

[6] S.Uehara, T.Hiroya, H. Kusanagi; K. Shigemura, H.Asada, “1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement”, i*n Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX*, Vol. 6803, San Jose, USA, January 2008

[7] S. Jumisko-Pyykkö and J. Häkkinen, “Evaluation of subjective video quality of mobile devices”, in *MULTIMEDIA ’05: Proc. 13th ACM international conf. on Multimedia*. New York, NY, USA: ACM Press, pp. 535–538, 2005.
[8] J. Häkkinen, M. Liinasuo, J. Takatalo, and G. Nyman, “Visual comfort with mobile stereoscopic gaming”, *Proceedings of SPIE*, vol. 6055, p. 60550A, 2006.
[9] M. McCauley and T. Sharkey, “Cybersickness: Perception of Self-Motion in Virtual Environments” in *Presence: Teleoperators and Virtual Environments*, 1(3), 311-318., 1992
[10] L. Meesters, W. IJsselsteijn, P. Seuntiëns, “A survey of perceptual evaluations and requirements of three-dimensional TV,” *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, No. 3, 2004, pp. 381 – 391.
[11] W. IJsselsteijn, P. Seuntiens and L. Meesters, “Human factors of 3D displays”, in (Schreer, Kauff, Sikora, edts.) 3D Video Communication, Wiley, 2005.
[12] Wandell, B.A., *Foundations of vision*, Sinauer Associates, Inc, Sunderland, Massachusetts, USA, 1995.
[13] D. Chandler, “Visual Perception (Introductory Notes for Media Theory Students)”, MSC portal site, University of Wales, Aberystwyth, available at http://www.aber.ac.uk/media/sections/image05.html
[14] IP. Howard and BJ Rogers, “Binocular Vision and Stereopsis”, in Oxford Univ. Press, NY, Oxford, 1995
[15] M. Wexler and J. Boxtel, “Depth perception by the active observer“, *Trends in Cognitive Sciences, 9,* pp. 431-438, Sept, 2005
[16] Julesz, B. *Foundations of Cyclopean Perception*, The University of Chicago Press, Chicago, 1971.
[17] A.Boev, A. Gotchev, K. Egiazarian, A. Aksay and G. Akar, “Towards compound stereo-video quality metric: a specific encoder-based framework”. *Proc. of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI 2006)*, Denver, CO, USA, 2006.
[18] M. Yuen, “Coding Artifacts and Visual Distortions”, in (*H. Wu. K. Rao eds), Digital Video Image Quality and Perceptual Coding,* ISBN 9780824727772 , CRC Press, 2005
[19] Merriam-Webster’s online dictionary, available at http://www.merriam-webster.com/dictionary/artifact
[20] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A.Koz, “Coding Algorithms for 3DTV—A Survey,” *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1606-1621, Nov. 2007
[21] A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C. Erdem, C. Weigel, A., “Scene Representation Technologies for 3DTV—A Survey,” *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1587-1605, Nov. 2007
[22] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information”, Doc. N8768, Marrakech, Morocco, January 2007.
[23] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data”, Doc. N8799, Marrakech, Morocco, January 2007.
[24] Lin, C., Ke, C., Shieh, C., and Chilamkurti, N. K. 2006. The Packet Loss Effect on MPEG Video Transmission in Wireless Networks. In *Proc. 20th international Conference on Advanced information Networking and Applications - Volume 1 (Aina'06) - Volume 01* (April 18 - 20, 2006). AINA. IEEE Computer Society, Washington, DC, 565-572.
[25] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, C. von Kopylow, "A Survey of 3DTV Displays: Techniques and Technologies," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol.17, no.11, pp.1647-1658, Nov. 2007
[26] Z. Wang, , A. Bovik, H. Sheikh and E. Simoncelli, “Image quality assessment: From error visibility to structural similarity”, *IEEE Trans. Image Processing*, vol. 13, No. 4, 2004, pp. 600-612
[27] J. Poikonen, J. Paavola, “Error Models for the Transport Stream Packet Channel in the DVB-H Link Layer”, P*roc. ICC 2006*, Istanbul, Turkey, 2006
[28] A. Boev, D. Hollosi, and A. Gotchev, “Classification of stereoscopic artefacts”, MOBILE3DTV Project report, available on http://mobile3dtv.eu
[29] A. Boev, D. Hollosi, and A. Gotchev, “Software for simulation of artifacts and database of impaired videos”, MOBILE3DTV Project report, available on http://mobile3dtv.eu

[P09] A. Boev, K. Raunio, M. Georgiev, A. Gotchev, K. Egiazarian, "Opengl-Based Control of Semi-Active 3D Display," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008* , vol., no., pp.125-128, 28-30 May 2008, doi: 10.1109/3DTV.2008.4547824

This page left blank

OPENGL-BASED CONTROL OF SEMI-ACTIVE 3D DISPLAY

ABSTRACT

We present a system for 3D visualisation, which combines "user-tracking" approach, used by displays with steerable optics, with generation of multiple views, typical for displays with fixed optical filter. Instead of eye-tracking, typical for the "user-tracking" approach, we propose a less computationally demanding head tracking, based on face detection. We investigate if the precise delivery of different images to each eye of the observer can be handled by the fixed optics of a multiview 3D display, and if continuous head parallax can be achieved.

Index Terms— 3DTV, multiview, auto-stereoscopic displays, GPU, OpenGL, 3D visualization, semi-active 3D display

1. INTRODUCTION

Until recently, it was common that observers of any 3D presentation were required to wear specially designed glasses. The next generation of displays which recently started to gain popularity, can create 3D representation of a scene without the need of glasses. These are known as autostereoscopic displays [1], [2], [3]. There are a number of taxonomies of 3D displays. A general one divides them into three basic types: holographic, volumetric and multiple-image screens [1], [2]. There are two types of multiple-image screens. The first type works by tracking the observer's eyes, and utilizes steerable optics to beam different images towards each eye. The second type uses fixed optics, and beams a number of different images (called "*views*") in different directions; the directions are selected in such way, that the eyes of an observer standing in front of the screen perceive different images. In [3], S. Pastoor classifies these two types as creating *eye-gaze-related image* and *fixed-plane image* correspondingly. Surman et al. use different taxonomy in [2] – the displays with steerable optics are named "*head position tracking displays*", while the ones with fixed optics are designated simply as "*multiview displays"*. In our study, we separate multiple-image displays into two general groups in respect to the optics - "*active*", which use eye-tracking and steerable optics, and "*passive*", which generate multiple images by means of fixed optics. We propose "*semi-active*" solution for a 3D display which combines active tracking with a passive, multiview display. When used by single observer, combination of less-demanding head tracking and partial view reorganization allows extending of the observation zone. If more than one user is present, the display operates in passive mode, allowing multiple observers to perceive 3D scene.

In the next section, we discuss the operation principles of autostereoscopic displays, the differences between "*active*" and "*passive*" autostereoscopic displays, as well as advantages and disadvantages of each approach. In Section 3, we present the general concept and the algorithm behind our "*semi-active*" 3D display system. Section 4 presents the algorithm used for face tracking, and Section 5 presents GPU-assisted visualization routine. The test setup and experiment results are presented in Section 6.

2. AUTOSTEREOSCOPIC DISPLAYS

Most of modern multiview displays use TFT screens for image formation [2-7]. The light generated by the TFT is separated into multiple directions by the means of special layer additionally mounted on the screen surface, as shown in Fig. 1a. Such layer is called "*optical layer*" [4], "*lens plate*" [3] and "*optical filter*" [9]. In this study we use the latter term. TFT displays do not have full-colour pixels. They recreate the full colour range by emitting light though red, green and blue coloured components (*sub-pixels*), usually arranged in repetitive vertical stripes. The optical filter mounted on top of the screen also has repetitive structure, which redirects the light passing through it. The intensity of the light rays passing through the filter changes as a function of the angle, as if the light is directionally projected [4]. Since sub-pixels appear displaced in respect to the optical filter, their light is redirected towards different positions, as shown in Fig. 1a. The image, seen on the screen from a particular direction is said to form a *view* [3], [4], [9]. As a result, differently coloured components of one pixel belong to different views. Respectively, the image formed by one view will be combination of colour components (sub-pixels) of various pixels across the TFT screen. When red, green and blue sub-pixels are visible from the same direction and appear

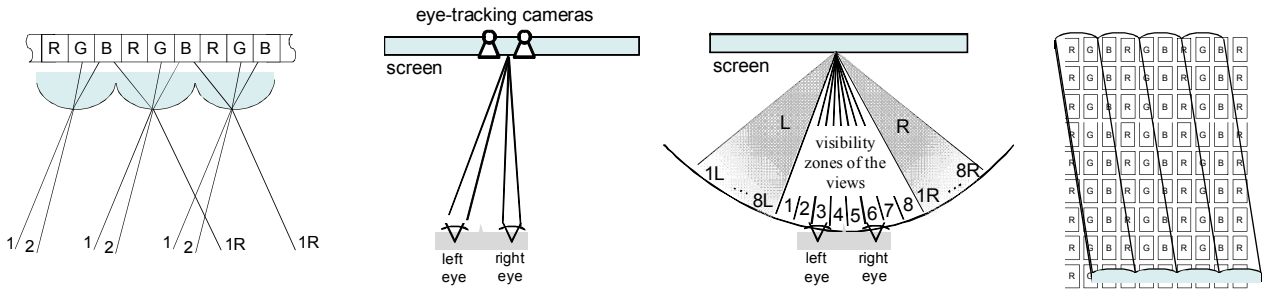


Figure 1. Autostereoscopic displays: a) Optical filter, separating the image into multiple views; b) "active" 3D display with steerable optics; c) "passive" multiview 3D display; d) slanted optical filter

close to each other, the triplet is perceived as one pixel. Such pixel is a building block of the view seen from that direction, and is sometimes referred to as "*poxel*" [10]. For every poxel there is a certain angle, from which it is perceived with maximal brightness – that angle we call *optimal observation angle* for the poxel. The vector, which starts from the poxel, and follows the optimal observation angle, is the *optimal observation vector* for the poxel.

Presently, the majority of autostereoscopic displays follow one of the following two approaches – using active optics, where two images are beamed in precise directions; and using fixed optics, where large number of images is beamed in fixed directions, across the observation zone of the display [2]. The "*active*" 3D displays use steerable optical filter and can accommodate to the head movement of the user by continuously readjusting the position of the filter in respect to the TFT screen. Such displays use head-tracking in order to point images precisely to the eyes of the observer. The early models used invasive tracking, requiring the user to wear optical marker or electronic transmitter. Modern displays with steerable optics use tracking cameras and eye-tracking software to adjust the position of the views [2], [11]. At least two cameras are used, to allow estimation of the position of each eye, as shown in Fig.1b. For the proper operation of the display, it is essential that the combination of software and hardware works in real-time. Typically, such displays create only two views, and are meant for a single observer [11], but initial steps are done towards development of a multiuser eye-tracking 3D display [2].

In contrary, a multiview 3D display has fixed optics, which creates a many views. For a large number of head positions, the eyes of an observer fall into the visibility zones of different views as exemplified in Fig. 1c. As a result, both eyes can perceive a scene at different angles, which enables 3D perception without wearing glasses. There are two common types of optical filters – lenticular sheet [8] which works by refracting the light, and parallax barrier which works by blocking the light in certain directions. The optimal observation vectors for all poxels of the same view are designed to intersect in a tight spot in front of the multiview display. From this spot, the view will be perceived with its maximal brightness, and we denote that spot as being the *optimal observation spot* of the view. Outside of the optimal observation spot, there is a range of observation angles, from which a given view is still visible, even though with diminished brightness. We refer that range to as the *visibility zone* of a view. In order smooth transition between the visibility zones to be created, and to balance the horizontal vs. vertical resolution of a view, the optical element is often placed at a slant over a standard LCD screen, as shown on Fig. 1d. This creates a specific correspondence of the pixels which belong to a certain view and the addressable sub-pixels of the display. In order to visualize multiple images on a multiview display, the images should be combined and their pixels reordered, following the configuration of sub-pixels belonging to each view. Such process is called "interdigitation" or sometimes "interzigging" [3], [12].

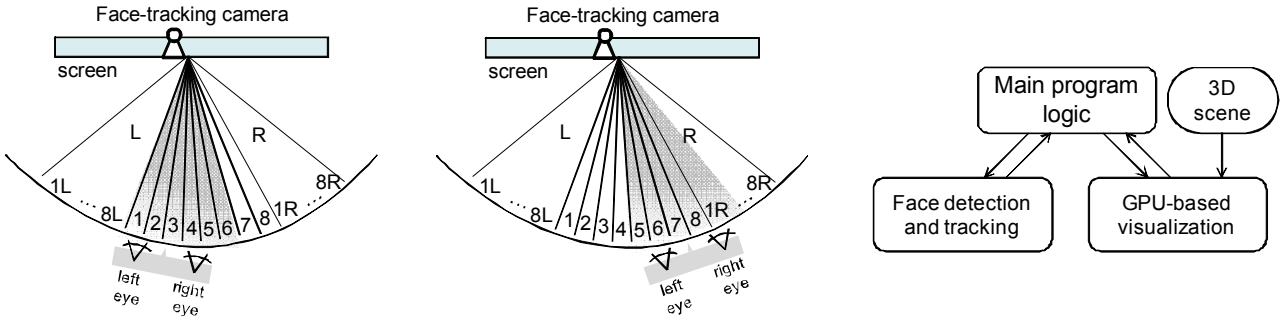
As the stereoscopic depth cues are perceived mostly in horizontal direction, the views on a typical multiview screen are ordered horizontally as well. When moving from left to right in front of the screen, the observation point will fall into the visibility zone of each view in a consecutive order, as shown in Fig. 1c. When the observation point moves past the visibility zone of the last view, the first view comes into visibility again, due to the regular structure of the lenticular sheet. Figure 1a shows two observation angles – 1 and 1', from which the same set of subpixels is visible. As a result, the full set of views that is seen in front of the screen is repeated to the sides.

Within given amount of sub-pixels, there is a trade-off between number of views and spatial resolution of a view. As a result, multiview displays offer lower resolution than the ones with steerable optic.

Even though a multiview display can be observed from wide range of angles, the view repetition limits the freedom to observe a scene from various angles. As an advantages, a multiview display works equally well for single or multiple observers, and does not require computationally demanding real-time eye-tracking.

3. SEMI-ACTIVE 3D VISUALISATION APPROACH

As an attempt to combine the ability to present a scene from many different angles of an "*active*" 3D display, with the lower computational requirements of a "*passive*" one, we derive an intermediate, "*semi-active*" approach for 3D visualization. We



a) b) c)

Figure 2. Selective view updating for continuous parallax: a) first position of the user's head; b) second position of the user's head, and c) block diagram of the active 3d visualization algorithm.

suggest that a combination of a multiview display, single camera and less-precise head-tracking is used. The software part of the system takes care that the observer's head is "surrounded" by a group of properly rendered views. Once the approximate position of the observer's head is found, the precise delivery of different images to the eyes is handled by the (passive) multiview optics.

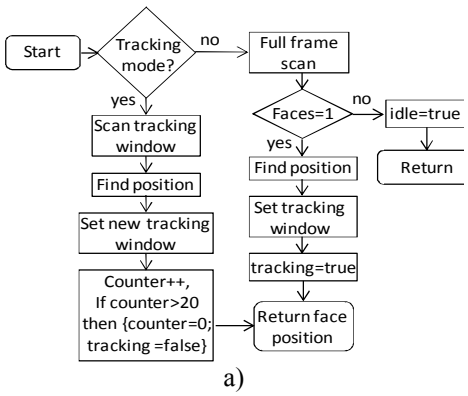
As shown in Fig. 1b, each view is seen from a number of observation spots, and the whole set of visibility zones is repeated on the sides, as depicted in Fig. 1c. If an observer moves laterally in front of the screen, after the visibility zone of the last view the first view comes in visible again, producing a characteristic "jump" of the 3D image [7]. However, we can provide a continuous parallax by replacing the views which are not visible with observations of the same 3D scene from new angles. For example, when the user's head is positioned as seen in Fig. 2a, the active views are from 1 to 6, and views 1 and 5 are seen by the left and right eyes correspondingly. When the user moves to the position shown in Fig. 2b, views 5 and 6 show the 3D scene at the same angles as before, and view 1 is updated to show the scene at a new angle. In reality, the eyes of the user fall into neighbouring views, and the view update happens well outside of the eye position. The head tracking has only to ensure the head of the observer is approximately at the centre of the set of updated views. Unlike the "*active*" eye-tracking approach, estimation of the distance between the observer and the display is not needed, as a set of properly rendered views can provide proper parallax to the eyes in a wide range of head positions. Also, real-time performance of the system is not necessarily critical, as the user is always "surrounded" by a safe margin of properly rendered views.

However, such approach is hard to be extended for multiple observers. If the eyes of different users fall into two of the observations zones of the same view, the system cannot render observation of the scene to satisfy both observers. To cope with this problem, we use two modes of operation. When two or more observers are detected, the system switches into "idle" mode and the display operates as a passive multiview 3D display. If only one user is detected, the system switches into "tracking" mode, providing wide observation parallax to the single observer.

Our software realization has a main routine which uses two modules, as seen from the block diagram in Fig. 2c. One module is responsible for face detection and tracking. Every time it is invoked, it returns either face position of a single observer or a flag, indicating that more observers are present. The second module is responsible for the GPU-assisted 3D visualization. Initially, it loads a scene description in the memory of the graphical accelerator. Each time a view update is needed, the main routine passes an observation angle to the GPU module. The module renders the scene at the required observation angles, and performs the interdigitation needed to replace the corresponding views.

4. FACE DETECTION AND TRACKING

The face detection module has two aims – to detect the number of observers and, in the case of single observer, to track continuously the position of the face in relation to the screen. It has two modes of operation, "tracking" and "full frame scan" as presented in the block diagram in Fig.3a. In order to optimize the performance, the module works mostly in "tracking mode" and performs face detection only inside a "tracking window" as shown in Fig.3b. Every time a face position is estimated, the position of the tracking window is updated, and



a) b)

Figure 3. Face detection module: a) block diagram of the module and b) position of

that the subsequent searches are done in a close neighbourhood of the location where face was previously found. On every 20th pass, a "full frame scan" is invoked, in order to check for additional observers. If more than one face is found, or no faces are found at all, the module returns special "idle "value, and the whole system switches to "idle" state as explained in the previous section. While in passive more, a "full frame scan" is periodically performed and when exactly one face is found, the system returns to "active" state.

The face detection module applies a two-stage hybrid technique. First, image areas having colour close to skin colour are detected and candidate face areas are determined. Second, feature-based face detection is performed in a sliding-window mode for the candidate areas only.

The skin detection algorithm utilizes two histogram colour models for the skin and non-skin respectively [13]. The histograms have been calculated using training skin and non-skin images in HSV colour space, for the chrominance channels (2-D histograms). A maximum likelihood ratio threshold is used to classify the processed pixel as skin colour pixel or non-skin colour pixels [14]. Connectivity analysis is carried out to eliminate background pixels and to unite skin-colour pixels. Thus, face candidate areas are formed and the subsequent face detection is run for these areas only.

The second stage is a feature-based face detection, which operates on the luminance channel of the colour image. It is a modification of the adaptive boosting algorithm [15], used recently by Viola and Jones [16] for simultaneously finding the best set of significant features of the pattern of interest (the face) and training a suitable classifier for that pattern. In our modification, optimal atomic decompositions are selected from various dictionaries of anisotropic wavelet packets to provide an adequate feature extraction [17]. Then, the adaptive boosting algorithm [15] is applied for finding the optimal subset of atoms. In contrast to the original Voila and Jones' threshold-type of weak learner, we employ a Bayesian-type of weak learner. It leads to a final strong classifier being able to place non-convex and even non-closed decision boundaries [17].

The cascade combination of skin-colour detection and Adaboost type of classification makes the whole system very fast and reliable.

5. GPU-BASED VISUALIZATION

The visualization module handles the rendering, interdigitation and display of a 3D scene. Initially, a group of 3D objects described using OpenGL is loaded to the memory of the graphical accelerator. When a scene update is needed, the main routine passes a value to the module, which indicates the approximate angle at which the used is looking at the screen. Since our screen is uses eight views, the first step of the visualization module prepare eight observations of the scene around the observation angle of the user. In our case we prepare 4 observations to the left and 3 to the right of the observer. For example, if the observer is approximately in the visibility zone of view 6, views 2, 3, 4 and 5 should contain observations from virtual cameras "placed" to the left of the user, as shown in Fig. 4. Accordingly, views 7, 8 and 1 should hold observations from virtual cameras to the right of the user. All eight observations are rendered and stored in 8 off-screen buffers the graphics memory. Usually, only 2 or 4 observations need to be updated, and the images stored in the rest of the buffers is reused.

Once all buffers are holding the needed observations, their contents are interdigitized and visualized on the screen. First, a special masking texture is prepared. The texture has some of its pixels transparent, and acts as a filter, which "passes through" only the pixels at certain position and colour, and renders everything else black. Since the masks needed for all views are shifted versions of the same structure, the same texture is positioned over the images in each buffer, but using the corresponding displacements. The rendered observation and the texture are blended together using "glBlendFunc" OpenGL function. The blended observation is sent to the accumulation buffer using glAccum. As unneeded sub-pixels are rendered black, and the masks used for different views do not overlap each other, adding new image to the accumulation buffer will only update the current view, without changing the others. Finally, "glAccumReturn" and "glutSwapBuffers" are used to send the interdigitized observations to the screen.

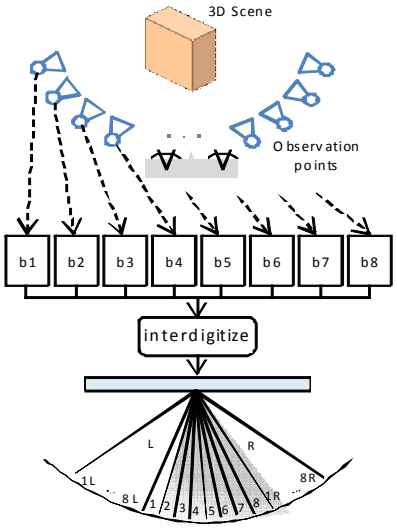
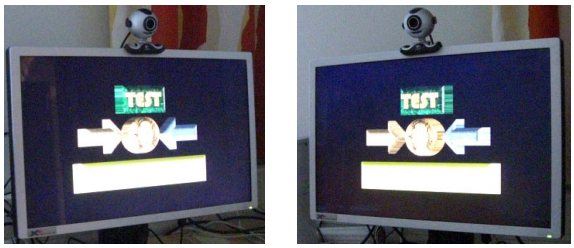


Figure 4. Visualization module

6. EXPERIMENTS

First, we calibrated the correspondence between the position of the face of the user in the image captured by the camera and the observation angle at which the user sees the display. We prepared eight test images, in which the subpixels belonging to a certain view were turned on maximal brightness; all other subpixels were turned off. With that image on the screen, there is a precise point, from which the screen is seen fully lit. This point is the optimal observation point of the current view. We ask the user to position his head in the optimal observation point in each view and record the horizontal position of the face as captured by the camera. Because the system works with approximate positions, the calibration works for any user of the display.



a) b)

Figure 5. Scene, rendered on a semi-active 3D display, as seen from different observation angles: a) left and b) right

Once calibrated, our system is able to provide much wider observation angle for a 3D scene than usual multiview display. On a modern computer equipped with GeForce 8800 graphical accelerator, the system is fast enough to deliver seamless experience and to work transparently for the user. Figure 5 presents a 3D scene as seen on our display from two different observation angles.

Currently our system visualizes only static 3D scenery. Future work will study the possibility of rendering 3D animation, as well as the description format suitable for rendering 3D animation on a "*semi-active*" 3D display.

7. CONCLUSIONS

We have proposed a "*semi-active*", GPU-based 3D visualization approach which combines the active tracking, characteristic for 3D displays with steerable optics, with a "*passive*", multiview 3D display. Instead of eye tracking, by using less-computationally demanding head tracking we are able to visualize a 3D scene from a wide range of observation angles. With moderate computational requirements our system is able to deliver a seamless experience to the end user.

Our system works in two modes of operation. When a single user is detected it operates in "tracking" mode, enabling continuous head parallax. If more than one observer is present, the system switches to "idle" mode, providing narrower freedom of movement, but satisfactory 3D visualization to multiple users.

ACKNOWLEDGEMENT

This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV. We thank Isabela Serano, Pauli Tuomola and Vladislav Uzunov for providing their source codes for face detection.

REFERENCES

[1] L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanlar and J. Watson: "An Assessment of 3DTV Technologies," *NAB Broadcast Engineering Conference Proceedings 2006*, pp. 456-467, Las Vegas, USA, April 2006.

[2] P. Surman, K. Hopf, I. Sexton, W.K. Lee, R. Bates, "Solving the 3D problem - The history and development of viable domestic 3-dimensional video displays", In (Haldun M. Ozaktas, Levent Onural, Eds.), *Three-Dimensional Television: Capture, Transmission, and Display* (ch. 13), Springer Verlag, 2007

[3] S. Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.

[4] C. Van Berkel and J. Clarke, "Characterisation and optimisation of 3D-LCD module design", in Proc. SPIE Vol. 2653, *Stereoscopic Displays and Virtual Reality Systems IV*, (Fisher, Merritt, Bolas, edts.), p. 179-186, May 1997

[5] C. van Berkel, D. Parker and A. Franklin, "Multiview 3D LCD," in Proc. SPIE Vol. 3012, *Stereoscopic Displays and Virtual Reality Systems III*, (Fisher, Merritt, Bolas, edts.), p. 32-39, 1996

[6] W. IJzerman et al., "Design of 2d/3d switchable displays," in *Proc. of the SID*, volume 36, Issue 1, pp. 98-101, May 2005

[7] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in *Proc. SPIE Photonics West 2002: Electronic Imaging*, vol. 4660, pp. 212-221, 2002

[8] Van Berkel, "Lenticular screen adaptor", US Pat. 6801243, issued Oct. 5, 2004

[9] W. Tzschoppe, T. Brueggert, M. Klipstein, I. Relke and U. Hofmann, "Arrangement for two-or-three-dimensional display", US pat. 2006/0192908, issued Aug. 31, 2006

[10] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science, vol. 194*, pp. 283-287, 1976.

[11] Phil Surman; Ian Sexton; Klaus Hopf; Wing Kai Lee; Richard Bates; Wijnand Ijsselsteijn; Edward Buckley, "Head Tracked Single and Multi-user Autostereoscopic displays," *Visual Media Production*, 2006. CVMP 2006. 3rd European Conference on , vol., no., pp.144-152, 2006

[12] J. Konrad and P. Agniel, "Artifact reduction in lenticular multiscopic 3-D displays by means of anti-alias filtering," *in Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 5006, pp. 336-347, Jan. 2003

[13] M.J. Jones and J.M. Rehg, 'Statistical color models with application to skin detection', International Journal of Computer Vision, vol. 46, No. 1, 2002, pp. 81-96.

[14] I. Serano and P. Tuomola, Skin color based image pre-processing, Tech. report, Tampere University of Technology, January 2006.

[15] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of Online Learning and an Application to Boosting", *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.

[16] P.Viola and M. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[17] V. Uzunov, A. Gotchev, K. Egiazarian, J.Astola 'Face Detection by Optimal Atomic Decomposition', *Proceedings of the SPIE*, Volume 5916, pp. 160-171 (2005).

This page left blank

[P10] A. Boev, K. Raunio, A. Gotchev and K. Egiazarian, “GPU-based algorithms for optimized visualization and crosstalk mitigation on a multiview display”, *Stereoscopic Displays and Applications XIX, Proc. SPIE 6803*, 68030K (2008), DOI:10.1117/12.761785

This page left blank

GPU-based algorithms for optimized visualization and crosstalk mitigation on a multiview display

Atanas Boev, Kalle Raunio, Atanas Gotchev, Karen Egiazarian
Institute of Signal Processing, Tampere University of Technology, Tampere, Finland

ABSTRACT

In this contribution, we present two GPU-optimized algorithms for displaying the frames of 2D-plus-Z stream on a multiview 3D display. We aim at mitigating the cross-talk artifacts, which are inherent for such displays. In our approach, a 3D mesh is generated using the given depth map, then textured by the given 2D scene and properly interdigitized on the screen. We make use of the GPU built-in libraries to perform these operations in a fast manner. To reduce the global crosstalk presence, we investigate two approaches. In the first approach, the 2D image is appropriately smoothed before texturing. The smoothing is done in horizontal direction by a 1-D filter bank driven by the given depth map. Such smoothing provides the needed anti-aliasing at the same filtering step. In the second approach, we introduce a higher number of properly blended virtual views than the display views supported and demonstrate that this is equivalent to a smoothing operation. We provide experimental results and discuss the performance and computational complexity of the two approaches. While the first approach is more appropriate for higher-resolution displays equipped with newer graphical accelerators, the latter approach is more general and suitable for lower-resolution displays and wider range of graphic accelerators.

Keywords: multiview display, crosstalk mitigation, GPU, visualization, 3D rendering

1. INTRODUCTION

Not very long ago, the spectators of a 3D visual presentation were usually required to wear porposely designed glasses, in order to perceive the scene in 3D. Recently, advances in display technology allowed the mass-production of screens, which could recreate a 3D scene without the need of glasses. Such displays are also known as "autostereoscopic", as initially they provided "left" and "right" images, separately targeted at the corresponding eye of the observer. The later generation of autostereoscopic displays is able to reconstruct multiple images of a scene, each seen from different observation angle. These are known as "multiview autostereoscopic" displays, and their advantage is that they can provide a 3D image to many users simultaneously, without requiring them to stay at a particular "sweet spot". Overview of various types of multiview displays can be found in[1,2]. It is expected that multiview displays utilizing lenticular sheets or parallax barrier will provide the first generation of 3D displays for widespread use[2].

Key factor for the wide adoption of 3D displays is the availability of compatible 3D content. An effective 3D scene representation format would need to support a large variety of 3D content creation and 3D visualization methodologies[3]. While there are many different formats for encoding 3D video, they can be divided in three main groups: *multiview video*, where two or more video streams showing the same scene from different viewpoints; *Video-plus-depth*, where to each pixel is augmented with information of its distance from the camera; and *dynamic 3D meshes*, where 3D video represented by dynamic 3D surface geometry[4]. Video-plus-depth format is suitable for multiview displays, as it can be used regardless of the number of views a particular screen provides[5,6]. Furthermore, video-plus-depth can be efficiently compressed[5]. Recently, MPEG specified a container format for video-plus-depth data, known as MPEG-3 Part 3[7,8]. On the downside, video-plus-depth rendering requires interpolation of occluded areas, which may be source of artifacts. This is being addressed by using layered depth images (LDI)[3] or by multi-video-plus-depth encoding[9].

A straightforward way to represent video-plus-depth is to encode the depth map as a gray scale picture, and place the 2D image and its depth map side-by-side. The intensity of each pixel from the depth map represents the depth of the corresponding pixel from the 2D image. Such format is sometimes referred to as 2D+Z, and a typical 2D frame looks like the one shown in Fig. 1. Due to its simplicity and versatility, we expect that the 2D+Z video format will be widely used with the first generation of multiview displays.

However, visualization of 2D+Z video on a multiview display requires additional computations. Based on the depth map provided with the scene, multiple observations should be rendered, and the pixels from these observations should be

Stereoscopic Displays and Applications XIX, edited by Andrew J. Woods, Nicolas S. Holliman, John O. Merritt,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6803, 68030K, © 2008 SPIE-IS&T · 0277-786X/08/$18

interleaved in the way required for the display. Furthermore, some multiview displays suffer from additional artifacts, which have to be corrected on-the-fly. It is likely, that a device with a 3D display would not only play video, but also will support gaming, or at least 3D menu navigation. With graphical accelerators being almost ubiquitous nowadays, we expect that many devices, equipped with multiview screens will also include a graphical accelerator (or GPU) of some kind. As OpenGL is the industry standard for programming GPUs, it is to be expected that such device is OpenGL-compatible too.

This paper studies how 2D+Z video can be rendered on a multiview display using OpenGL coping with cross-talk artifacts, inherent for such displays. In the next section, we discuss the principles of work of multiview displays, and the typical visual artifacts, created by them. In Section III we explain the reason for crosstalk, being the most severe artifact for the screen used in our experiments, and an approach to mitigate it. The following section describes two alternative algorithms for crosstalk mitigation that we implemented using OpenGL. Finally, we present the results of the two implementations, comparing speed, memory requirements and visual quality.

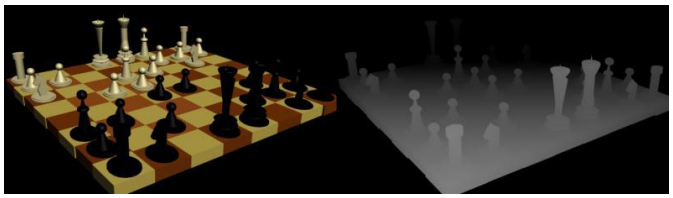
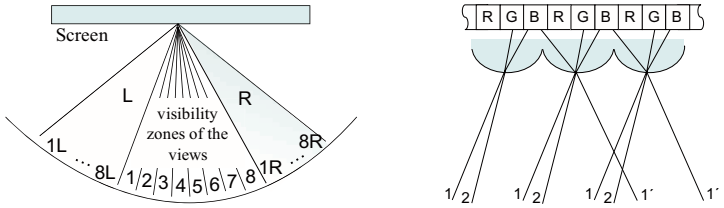


Fig.1. An example 2D+Z image

2. MULTIVIEW DISPLAYS

2.1 Principles of work

Multiview autostereoscopic display creates 3D illusion by "casting" different images in different directions. Currently, the majority of multiview displays are using TFT screen for image creation[10,11,12,13,14,15]. Additional optical layer is used to redirect the light passing through the LCD. As a result, only a subset of the pixel color components (also known as sub-pixels) are seen from certain observation angle. The set of sub-pixels visible from a certain angle forms an image also known as a "view". The area, from which a particular view can be seen, is called the "visibility zone" of that view. Since stereoscopic depth cues are perceived mostly in horizontal direction, visibility zones of the views on a typical multiview screen are ordered horizontally as shown in Fig. 2a. The angle between the observation zones is designed to allow the eyes of an observer staying in front of the screen to perceive different images. Due to the repetitive nature of both the TFT and the optical layer, the subpixels of each view are seen from more than one direction, as shown in Fig. 2b. As a result, when the observer moves past the visibility zone of the last view, the first view comes into visibility again.

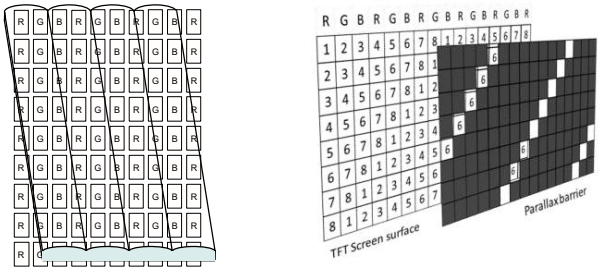


a) b)

Fig.2. Light redirection in multiview displays: a) visibility zones of the views, b) optical layer, redirecting the light of the sub-pixels

In order to visualize multiple images on a multiview display, the images should be combined and their pixels reordered, following the configuration of sub-pixels belonging to each view. Such process is called "interdigitation"[11,16], or sometimes "interzigging"[16]. When the images contain many observation of the same scene, and they are interdigitized properly, the observer is able to perceive the scene in 3D.

Early designs of multiview displays had discrete boundaries between the viewing zones[11,12]. This is the source of two common artifacts, found in autostereoscopic displays. One is "image flipping", caused by the noticeable transition between the viewing zones[12]. Another is "picket fence effect", also known as "banding" - a moiré-like artifact caused by the gaps between subpixels being magnified by the lenticular sheet[11]. In order to mitigate these effects, some vendors intentionally broaden the observation angle of the pixels[14], interspersing the viewing zones. It is also speculated, that blurring the boundaries between the viewing zones can increase the apparent number of views[13,21]. In 1996, van Berkel proposed an elegant solution to the two problems[13]. He suggested that a lenticular sheet could be placed at a slant over a standard LCD screen, as shown in Fig. 3a. This approach removes the picket fence effect, creates smooth transition between the views and at the same time balances the horizontal vs. vertical resolution of a view. Another solution with similar effects is "wavelength-selective filter array" proposed by 4D-Vision GmbH in[15]. Essentially, the filter is a slanted parallax barrier which covers the display and defines particular light penetration direction of each subpixel. Depending on the observation angle and the distance to the observer, most of the sub-pixels are masked. Only the sub-pixels which belong to one view are visible, as it is exemplified in Fig. 3b.

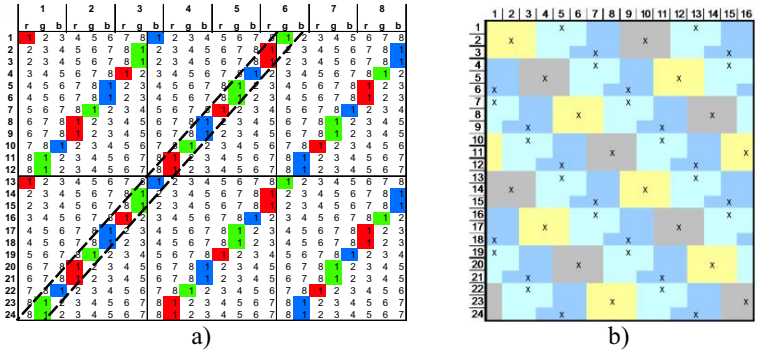


a) b)

Fig.3. Slanted optical layers: a) slanted lenticular sheet and b) slanted parallax barrier

2.2 Visual artifacts created by slanted optical layer

While both solutions – slanted lenticular sheet and slanted parallax barrier – help to reduce banding and image flipping artifacts, they also create problems on their own. Due to the slant, the sub-pixels of a certain view appear on a non-rectangular grid. Example configuration of sub-pixels forming one view is shown in Fig. 4a. Furthermore, sub-pixels of different colors do not appear horizontally adjacent in one view. An area, in which red, green and blue sub-pixels appear close to each other, is perceived as single, full-color element of that view. Such element is sometimes referred to as "poxel"[18]. The color components of the newly formed poxel are coming from different addressable pixels which depend on the topology of the view. They can be spatially approximated by using different areas of the image, often with non-rectangular shape. Figure 4b shows one possible separation of the screen into poxels, if the sub-pixel topology follows the one in Fig. 4a. The need to resample onto a non-rectangular grid of a view requires specially designed anti-aliasing filters. A methodology for design of such filters has been proposed and thoroughly studied [16, 19].



a) b)

Fig.4. Topology of a view when using slanted optical layer: a) subpixels which belong to one view and b) areas of the screen which contain all three sub-pixel colors. Both are fragments of a repetitive pattern which covers the screen surface.

The use of slanted optical layer is also responsible for another artifact, called "ghosting". As sub-pixels have rectangular shape, they appear displaced in respect to the center of the slant, as plotted with dashed line in Fig. 4a. Sub-pixels belonging to different rows appear with different horizontal shift under slant, and their visibility zones are slightly

different. Additionally, some vendors broaden the observation angle of a sub-pixel, in order to create more uniform view[14]. The visibility zones of the neighboring views become interspersed, and any observation spot falls into the viewing zones of different views, as shown in Fig. 5a.

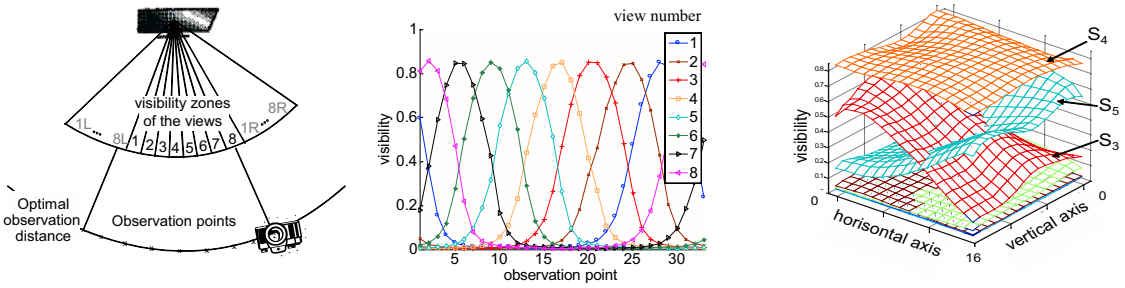
As a result, images which belong to many views are simultaneously visible, even with a single eye, which can be regarded as inter-channel crosstalk. The crosstalk manifests itself as multiple contours around object shapes, scattered in horizontal direction. A snapshot of a multiview screen exhibiting ghosting artifacts is shown in Fig. 5. While studying an 8-view 3D display, we found that ghosting is more pronounced and more annoying artifact than aliasing[20]. The presence of crosstalk, especially noticeable on the objects with pronounced depth, can completely destroy the ability to perceive these objects in 3D. Other authors also agree that ghosting artifacts hinder the perception of binocular depth cues[21].

a) b)

Fig.5. Ghosting artifacts: a) interspersed visibility zones and b) snapshot of a multiview screen, as example for ghosting artifacts

3. CROSSTALK MITIGATION

In a previous study, we proposed a methodology for assessing the crosstalk of an arbitrary screen[20]. The visibility of each view is measured at a number of observation points, placed along an arc, as it is shown in Fig. 6a. The measurements allow estimation of the individual contribution of each view to the crosstalk for various observation angles. The plots in Fig. 6a and 6b present results obtained measuring the crosstalk of an 8-view auto-stereoscopic screen manufactured by X3D Technologies GmbH. When moving the observation point along the arc, the views gradually come into and disappear from visibility, as is seen in Fig. 6b. The visibility peaks of each view appear at equal distances, and depending on the observation angle, various combinations of views with different intensities are seen on the screen. The surfaces in Fig. 6c represent the visibility of different views across the screen surface as measured at one of the observation points. As seen on the figure, as one view is predominantly seen, the crosstalk contribution of its neighbors changes both in horizontal and vertical directions.



a) b) c

Fig.6. Crosstalk measurements: a) observation points, b) visibility of each view across the observation points, c) visibility of each view across the screen surface, as measured at one observation point.

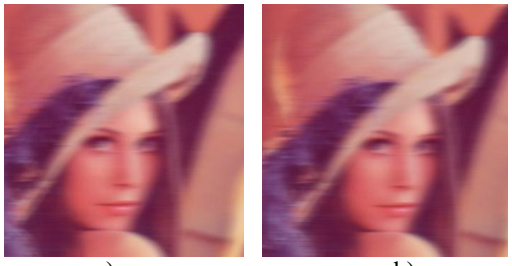
Based on the crosstalk measurement, we are able to simulate the images seen on X3D-23" display from various distances and observation angles. For example, let us assume that an observer looks at the screen from the typical observation distance of 150cm and has inter-ocular distance of 65mm. Then, the left eye sees predominantly the image in view number “3” with and overlaid images from views numbers “2” and “4” with brightness levels as shown in Fig. 7a.

Similarly, the picture seen from the right eye is dominated by view "4", with contributions from other views as seen in Fig. 7d. If the object on the screen is rendered at a certain depth, which results in 10 pixels disparity between the images in all the view, the image presented to the left eye would look similar to the one in Fig 7b, and the image presented to the right eye – as the one in Fig. 7c. The fusion of such stereoscopic pair suggests several possible depth levels, and as result is hard to be perceived at any certain depth.

a) b) c) d)

Fig.7. Crosstalk simulation: a) brightness of the views as perceived by the left eye, b) the picture, seen by the left eye, c) the picture seen by the right eye, and d) brightness of the views as perceived by the right eye

Konrad *et al* propose a pre-compensation algorithm for reducing the crosstalk in stereoscopic displays[22]. However, their approach is not suitable for multiview displays with slanted optical layer. For such case, pre-compensation mitigates the effect for a certain observation angle only, while amplifying it for other angles. As multiview display is intended for many observers, it is desirable to mitigate the ghosting artifacts for all observation angles simultaneously. In a properly formatted 3D scene, observations of any object have horizontal disparity, and ghosting artifacts appear in horizontal direction only. In order to mitigate the crosstalk, we propose smoothing all observations in horizontal direction, where the level of smoothing depends on the amount of the disparity. For a scene in 2D+Z format, this corresponds to smoothing of the 2D image, with level of smoothing depending on the absolute depth values of the pixels. The further away from the screen level an object is set to appear, the bigger disparity between its observations would be, and larger amount of smoothing would be required. For example, when the same images with the same disparity as in the previous example are pre-filtered with smoothing filter in horizontal direction, and then overlaid using the brightness levels from Figures 7a and 7d, the result looks as the stereo-pair in Fig. 8. Such pair is much easier to be fused by the observer, and results in a flat image, floating approximately 20cm in front of the screen surface.



a) b)

Fig.8. Horizontally smoothed images, undergoing the same crosstalk as in Fig. 7: a) the result as perceived by the left eye and b) the result as perceived by the right eye.

4. OPENGL BASED IMPLEMENTATION

4.1 Building 3D image from 2D+Z data

In order to display 2D+Z image on a multiview display, the 2D+Z data should be converted into multiple observations of a 3D scene. Typically, this is done in a three step process – first, given the angle of observation the disparity corresponding to different depth levels is calculated; then, the pixels in the 2D image are displaced horizontally according to the calculated disparity; finally, the pixels which belong to previously hidden parts of the image are recovered using interpolation. Recovery of hidden pixels is sometimes called "disocclusion" and various approaches exist – depth-level based[23], segmentation-based[24] and texture-based[25].

As 3D graphical accelerators are optimized for geometry-related calculations, we decided to utilize GPU as a convenient tool for 2D+Z to multi-view conversion. We build dynamic 3D mesh which is modeled by the depth map and uses the 2D scene as a texture. For each frame, an image of the mesh is rendered from as many observation points as needed for a multiview display (Fig. 9). The mesh is build following the "DMesh" algorithm, which is optimized for speed[26]. However, since for video the level of details changes from frame to frame, we had to use mesh with uniform resolution of one vertex per pixel. When a 3D accelerator renders a textured mesh, usually liner approximation is used, which causes specific "rubber sheet" artifacts[24,26]. For dealing with these artifacts we smooth the depth map before modeling the mesh, a technique proposed in[26].

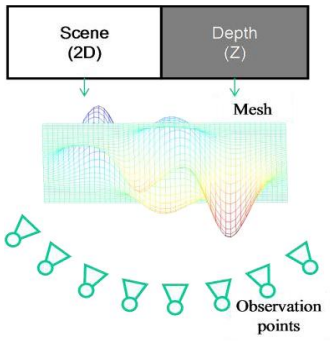


Fig.9. Rendering of multiple observations from 2D+Z image

Once rendered, each observation should be mapped to the sub-pixels which belong to the corresponding view. In order to save memory, our algorithm maps each observation to the final result as soon as it is rendered. The rendering and mapping operation is done in a common loop which performs the following steps:

- The textured mesh is rotated in respect to the "camera", in order to generate observation of the scene at a certain angle
- The image "seen" through the camera is rendered to an off-screen buffer. The rendering is done in "orthogonal" mode, which eliminates perspective scaling of the scene.
- Second texture is placed over the rendered view. The texture has some of its pixels transparent, and acts as a filter, which "passes through" only the pixels at certain position and color, and renders everything else black. Furthermore, as the masks for all views are shifted versions of the same structure, we use only one mask, and shift it each time a new observation arrives under it.
- The rendered observation and the texture are blended together using "glBlendFunc" OpenGL function. The blending acts as a color filter which leaves only the sub-pixels (certain color components of certain pixels) needed for the current view.
- The blended observation is sent to the accumulation buffer using glAccum. As unneeded sub-pixels are rendered black, and the masks used for different views do not overlap each other, adding new image to the accumulation buffer will only update the current view, without changing the others.

The steps above are repeated as many times as many views are needed for the 3D display. Finally, "glAccumReturn" and "glutSwapBuffers" are used to send the interdigitized observations to the screen.

As discussed, in order to mitigate the ghosting artifacts a smoothing operation which varies with the depth should be used. Depth of field (DoF) rendering, which aims at blurring the foreground and background, while leaving objects at a certain depth "in focus" is a similar problem, which is often solved using OpenGL primitives. The difference in our case is that we need to smooth the image only in horizontal direction, instead of both in horizontal and vertical direction as it is done for rendering DoF. Three widely-used approaches exist for DoF simulation using OpenGL – pre-filtering, scattering and point-based splatting. Pre-filtering works by decomposing the scene into sub-images with different depth levels, applying different blur filters, and then blending the sub-images together[27]. Scattering is an approach with blends together displaced semi-transparent versions of the same image, where displacement depends on the amount of blur needed[28]. If the scene is rotated around a central point, this results in small displacements of objects close to point of

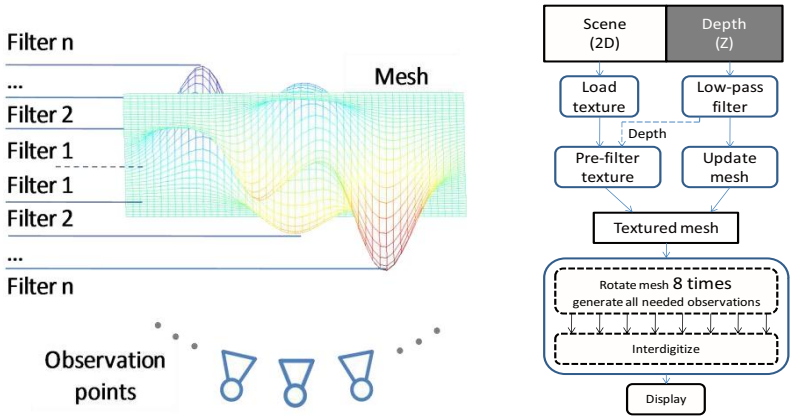
rotation, and large – for objects far from that point. Point-based splatting presents the object as a cloud of points without connectivity, and the amount of blur is achieved by controlling the diameter of the point[29]. The last approach produces superior quality, but is computationally intensive, and requires different scene representation than mesh. In the next two sections we describe two alternative approaches to crosstalk mitigation. The first is similar to pre-filtering, and the second is adapted version of the scattering technique.

4.2 Pre-filtering of the texture

Our first algorithm for crosstalk mitigation employs pre-filtering of the 2D image, before using it as a texture on the mesh. The aim is to apply different amount of smoothing for areas at different depth level. Depth values in the middle of the scale will result in objects appearing close to the screen level, and observations of such objects will have small or no disparity. In such areas crosstalk is not visible, only an anti-aliasing filter should be applied, as marked with "Filter 1" in Fig. 10a. The further away from the depth is from the middle of the scale, the further away the object appears from the screen level. Areas with such depth will experience pronounced ghosting artifacts, and should be filtered with more restrictive low-pass filters in horizontal direction.

The algorithm has the following steps, also shown in Fig. 10b:

- The 2D scene is loaded to the texture memory of the graphics card
- The depth map is loaded and filtered with low-pass filter in order to mitigate "rubber sheet" artifacts
- The filtered depth map is used to select the areas of the texture, and each area is filtered using different smoothing filter
- The mesh is updated to reflect the current depth map
- The texture is applies to the mesh
- Following the algorithm, described in the previous chapter, multiple observations of the mesh are created and interdigitized according to the view topology of the display. The display which we used in the tests required 8 observations



a) b)

Fig.10. Crosstalk mitigation with pre-filtering: a) filters used for different depth levels and b) block diagram of the algorithm.

The texture is filtered according to the depth levels in a separable manner as seen in Fig. 11a. We use eight filters for the whole range of depth values. The depth values are 8 bit, from 0 to 255, and the value 127 represents depth equal to the screen level. First, the texture is filtered along the columns by a 1D filter with impulse response h_0, which acts as a simple anti-aliasing filter. The result is separately filtered eight times along the rows, using 1D filters h_0 to h_7, resulting in eight images with various smoothing in horizontal direction. Eight masks are prepared, passing different range of depth values, according to the distance from the screen level (d=127), as shown in Table 1. Each mask is applied to the corresponding filtered image, and the result is blended together in the accumulation buffer. As a result, the areas with

depth close to the screen level are filtered with h_0 in horizontal and vertical direction, which serves as an anti-aliasing filter. The areas, which appear further away from the screen level, are filtered with h_0 in horizontal direction and other, more restrictive low-pass filter in vertical direction, which mitigates aliasing and ghosting artifacts at the same time. For filter implementation we use a low level, GPU-optimized library called CUDA[30].

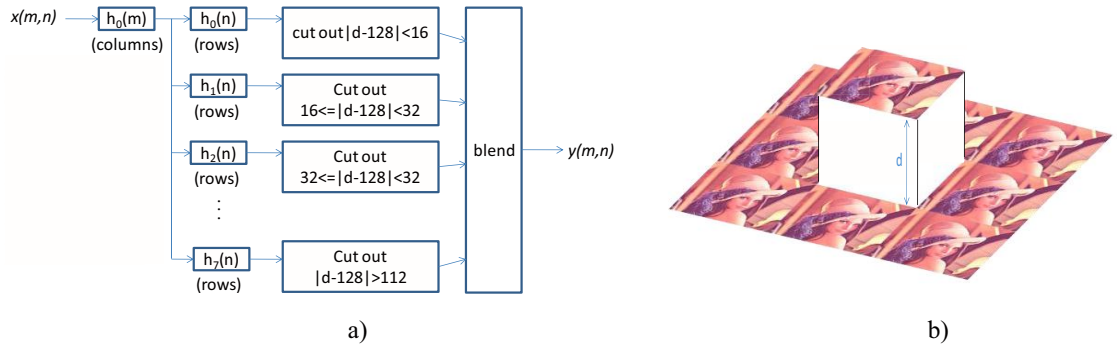


Fig.11. Variable smoothing according to the depth level: a) filter bank and area masking used in the algorithm and b) test setup for filter selection

When choosing the filter for a given depth level, we aimed at the shortest filter which would smooth the image just enough, that the ghosting artifacts at this level are not visible. As the ability of the human eye to see double edges depends on many factors, some of which hard to express in mathematical form, we performed a simple subjective test in order to select the optimal filters for our particular 3D display. We created impulse responses with a Gaussian shape with lengths from 1 to 50. The sigma of each Gaussian shape was selected such, that the resulting low-pass filter has the steepest slope for the given length. We prepared 15 test images, in which the same 2D image, made from tiled images of Lena. In the depth map, most of the image is at the screen level, except the central tile, which is rendered at different depths, as exemplified in Fig. 11. One test image has its central tile at the screen level, in seven it appears in front, and in other seven it appears behind the screen at distances as shown in Table 1. Each test image exhibited various amount of aliasing and ghosting artifacts. While observing the screen at the optimal observation distance of 150cm, we tried all 50 filters on the center tile. The shortest filter, which smoothed image just enough to mitigate the ghosting artifacts, was selected as the optimal for that depth range.

Table 1 – Selection of filters for different depth ranges

Filter	Distance from screen level, $\lvert d-128 \rvert$	Depth range	Filter length	Sigma
h_0	16	112..143	3	0.5
h_1	32	96..111, 144..159	5	0.84
h_2	48	80..95, 160..175	7	1.17
h_3	64	64..79, 176..191	11	1.83
h_4	80	48..63, 192..207	15	2.5
h_5	96	32..47,208..223	23	0.26
h_6	112	16..31, 224..239	33	3.84
h_7	128	0..15,240..255	47	7.83

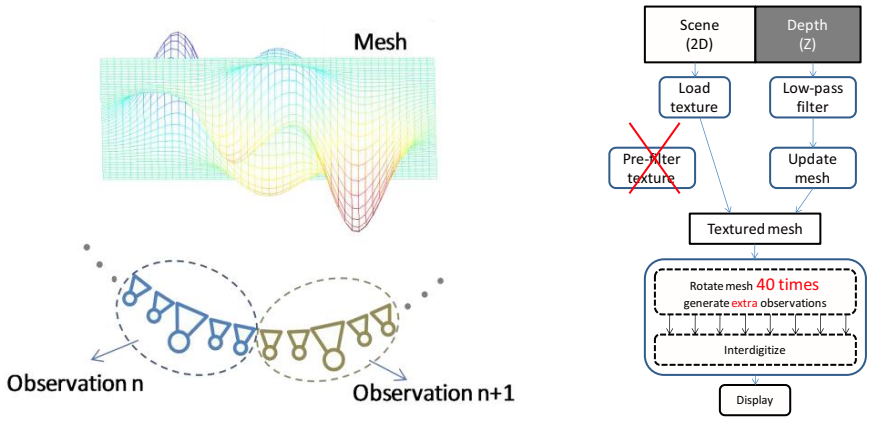
4.3 Using extra observation points

As an alternative approach, we adapt image scattering technique for crosstalk mitigation, by blending extra observations with the ones needed for the multiview display. Around each observation point used in previous approach, we place additional observation points at equal angles, grouped as shown in Fig. 12a. The angle between adjacent cameras from neighboring groups is the same as the angle between cameras inside a group. The images rendered from a group of observation points are blended together in a single image, which is mapped to the subpixels which belong to one view of

the screen. The algorithm follows the same steps as before, however the texture is not pre-filtered, and additional observations are rendered instead, as illustrated in Fig. 12b. We use 4 additional observations for each view. Since our screen requires 8 views of a scene, 40 observations are rendered, and are blended together in 8 groups of 5 observations. The rendering and mapping loop, described in Section 4.1 is modified and uses the following steps:

- The accumulation buffer is emptied.
- The textured mesh is rotated in respect to the camera. Now the angle of rotation is five times smaller than before.
- The image "seen" through the camera is rendered to an off-screen buffer in orthogonal mode.
- A masking texture is placed over the rendered view. The number of the observation is divided by 5, and the integer part of the result is used as the number of the mask. This ensures that for the mask for view 1is used for first 5 observations, the mask 2 is used for the next 5 observations, and so on.
- The rendered observation and the texture are blended together using "glBlendFunc".
- The blended observation is sent to the accumulation buffer using glAccum. However, now the transparency factor is 0.2. This ensures that the first five observations are blended together in the sub-pixels corresponding to view 1, the next 5 observations are blended together in view 2, and so on.

The steps above are repeated five times more than the number of views needed for the 3D display. Finally, "glAccumReturn" and "glutSwapBuffers" are used to send the interdigitized observations to the screen.



a) b)

Fig.12. Crosstalk mitigation with pre-filtering: a) position of the extra observation points in respect to the original ones and b) block diagram of the algorithm.

When an object appears at a given depth, its observations have disparity, corresponding to that depth. As many views are seen at the same time, the screen acts as if it is blending these observations, multiplying each one with a given factor. Let us assume the disparity between the observations of an object is 20 pixels, and the screen is observed at an angle for which the crosstalk coefficients are the same as in Fig. 7a. In that case, the screen acts as a filter which impulse response has 20 zeroes between each significant value, and the values are the same as the crosstalk coefficients measured for that observation angle. Such impulse response is plotted in Fig. 13a, and the frequency response of the corresponding filter is shown in Fig. 13b. As seen from the frequency response, middle and high frequency ripples are passed by the filter cause ghosting artifacts. An image, filtered with such a filter is shown in Fig. 14a. When five observations are rendered for each view, and the angle between them is five times smaller, the disparity between the images is five times smaller, too. If the brightness factor of each image is 0.2, the screen acts as a filter with impulse response as the one given in Fig. 13c. The corresponding frequency response is shown in Fig. 14d. An image, filtered with this impulse response (shown in Fig. 14b) looks smoother. Still, the image suffers from ghosting artifacts, as seen in the enlarged fragment in Fig. 14c. However, when the display is observed from the optimal distance of 150cm, the eye acts as additional low-pass filter, and mitigates the additional peaks in the frequency response in Fig. 13d. We experimented with various amount of

additional observations, using the setup from Fig. 11b. We found that 5 observations per view are enough to mitigate the ghosting artifacts for the maximum depth levels rendered by out algorithm.

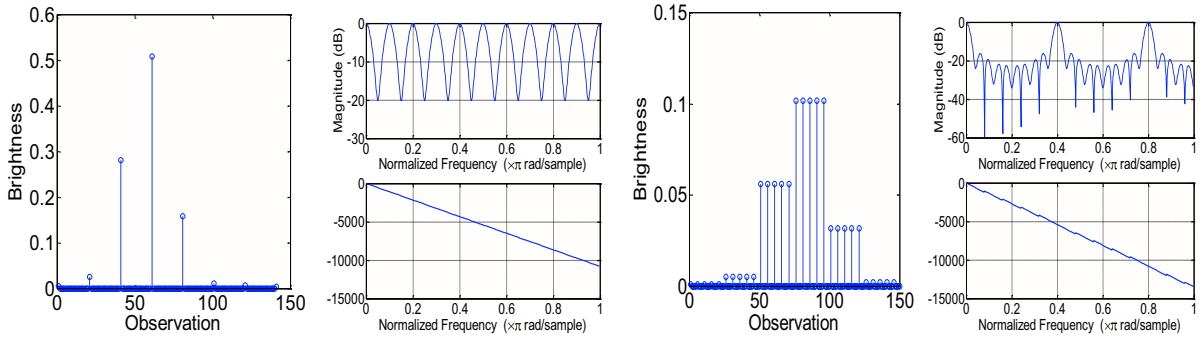


Fig. 13. Crosstalk regarded as a filter: a) combined impulse response of the view visibility and disparity of 20 pixels, b) frequency response of the filter with such impulse response, c) impulse response when extra observations are blended and d) frequency response when extra observations are blended.

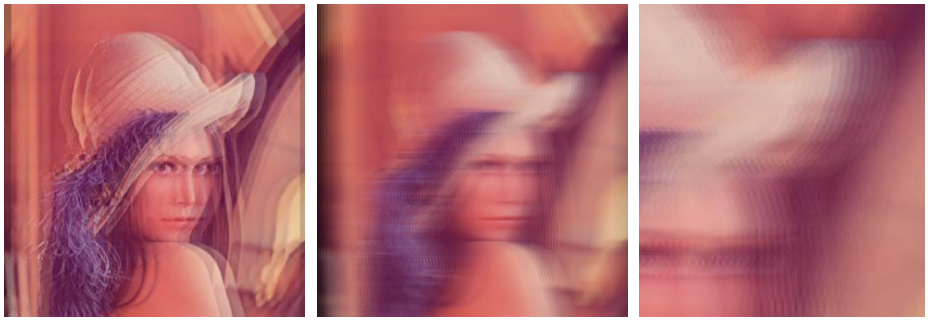


Fig. 14. Images filtered with the impulse responses from Fig. 13: a) crosstalk as a result from disparity of 20 pixels, b) result, when blending extra observations and c) enlarged fragment of b), emphasizing the "ringing" artifacts

5. RESULTS

5.1 Visual improvements

For visual comparison of the results, we present three snapshots of our display, showing various test images. The test setup from Fig. 11b is used, and the center tile is positioned at the maximum distance in front of the screen, allowed by our software. If no crosstalk mitigation is used, the tile exhibits strong ghosting artifacts, as seen in Fig. 15a. When looking at the scene with both eyes, it is impossible to perceive the central tile at any particular depth. If the texture is pre-filtered (Fig. 15b), the center tile looses details, but is immediately seen as floating approximately at 50cm in front of the screen. When using extra observations and without texture pre-filtering (Fig. 15c), the result is virtually indistinguishable from the pre-filtered version and yields satisfactory 3D effect.

5.2 Benchmark results

We measured the execution times of our algorithms on a GeForce 8800 GPU. Each stage was separately run 256 times, and the mean execution time for the main blocks of each algorithm is presented in Table 2. The overall speed in frames per second for various image and depth sizes of both approaches is also presented. The texture pre-filtering algorithm run faster, but it needs an optimized low-level filtering library, which works only with the latest generation of GPUs. The algorithm using extra observations produces comparative speeds using only high level OpenGL functions, and can be used with wider range of graphical accelerators. The most time-consuming stage for both approaches is the mesh update. We are investigating alternative approaches in mesh updating, which might execute faster.

a) b) c)

Fig. 15. Snapshots of the display: a) the test image from Fig. 14b rendered without crosstalk mitigation, b) the same test image, rendered with texture pre-filtering and c) the same test image, rendered with extra observations

Table 2 – Execution times and frames per second

	Algorithm 1: Texture pre-filtering			Algorithm 2: Using extra observations		
Frame size	1920x1200	720x400	640x400	1920x1200	720x400	640x400
Load texture	0.015sec	0.002sec	0.001sec	0.015sec	0.002sec	0.001sec
Update mesh	0.464sec	0.057sec	0.052sec	0.464sec	0.057sec	0.052sec
Filter texture	0.047sec	0.007sec	0.006sec	-	-	-
Render and interditigize	0.039sec	0.004sec	0.003sec	0.177sec	0.021sec	0.018sec
Total frames per second	1.77FPS	14.29FPS	16.13FPS	1.52FPS	12.50FPS	14.08FPS

6. CONCLUSION

We propose two algorithms for GPU-based perceptually optimized rendering of 2D+Z video frames on a multiview display. We identify crosstalk as a factor which prevents proper 3D perception, and suggest two alternative approaches for mitigating its effects. As a case study, we use crosstalk measurements of an 8-view display to optimize 2D+Z content for it. Both algorithms – using texture pre-filtering and using extra observations – improve the depth perception of rendered content. We present the time for execution of both algorithms, tested on a GeForce 8800 GPU. The algorithm using texture pre-filtering performs faster, while the algorithm using extra observations is applicable to wider range of graphical accelerators.

ACKNOWLEDGEMENT

This work is supported by EC within FP6 (Grant 511568 with the acronym 3DTV) and by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program (2006 - 2011)).

REFERENCES

[1] L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanlar and J. Watson, "An Assessment of 3DTV Technologies," *NAB Broadcast Engineering Conference Proceedings 2006*, pp. 456-467, Las Vegas, USA, April 2006.

[2] P. Surman, I. Sexton, R. Bates, W. K. Lee, K. Hopf, and T. Koukoulas: "Latest Developments in a Multi-User 3D Display," in *Proc. SPIE Vol. 6016, Three-Dimensional TV, Video, and Display IV* , 2005.

[3] A. Alatan, Y. Yemez, U. Gudukbay, X. Zabulis, K. Muller, C. Erdem, C. Weigel, A., "Scene Representation Technologies for 3DTV—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1587-1605, Nov. 2007

[4] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis, A.Koz, "Coding Algorithms for 3DTV—A Survey," *Circuits and Systems for Video Technology, IEEE Transactions on* , vol.17, no.11, pp.1606-1621, Nov. 2007

[5] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. IJsselsteijn, M. Pollefeys, L. Van Gool, E. Ofek, and I. Sexton, "An evolutionary and optimized approach on 3D-TV," *in Proc. Int. Broadcast Conf*., Amsterdam, The Netherlands, Sep. 2002, pp. 357–365.

[6] C. Fehn, "3D-TV using depth-image-based rendering (DIBR)," in *Proc. Picture Coding Symp*., San Francisco, CA, USA, Dec. 2004.
[7] Text of *ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information*, ISO/IEC JTC1/SC29/WG11, Jan. 2007, Doc. N8768, Marrakesh, Morocco.
[8] *Text of ISO/IEC 13818-1:2003/FDAM2 Carriage of Auxiliary Data*, ISO/IEC JTC1/SC29/WG11, Jan. 2007, Doc. N8799, Marrakech, Morocco.
[9] C. Fehn, N. Atzpadin, M. Muller, O. Schreer, A. Smolic, R. Tanger, P. Kauff, P., "An Advanced 3DTV Concept Providing Interoperability and Scalability for a Wide Range of Multi-Baseline Geometries," *Image Processing, 2006 IEEE International Conference on* , vol., no., pp.2961-2964, 8-11 Oct. 2006
[10] P. Surman, K. Hopf, I. Sexton, W.K. Lee, R. Bates, "Solving the 3D problem - The history and development of viable domestic 3-dimensional video displays", In (Haldun M. Ozaktas, Levent Onural, Eds.), *Three-Dimensional Television: Capture, Transmission, and Display* (ch. 13), Springer Verlag, 2007
[11] Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) *3D Video Communication*, Wiley, 2005.
[12] C. Van Berkel and J. Clarke, "Characterisation and optimisation of 3D-LCD module design", in *Proc. SPIE Vol. 2653, Stereoscopic Displays and Virtual Reality Systems IV*, (Fisher, Merritt, Bolas, edts.), p. 179-186, May 1997
[13] C. van Berkel, D. Parker and A. Franklin, "Multiview 3D LCD," in *Proc. SPIE Vol. 3012, Stereoscopic Displays and Virtual Reality Systems III*, (Fisher, Merritt, Bolas, edts.), p. 32-39, 1996
[14] W. IJzerman et al., "Design of 2d/3d switchable displays," in *Proc of the SID, v*olume 36, Issue 1, pp. 98-101, May 2005
[15] A. Schmidt and A. Grasnick, "Multi-viewpoint autostereoscopic displays from 4D-vision", in *Proc. SPIE Photonics West 2002: Electronic Imaging*, vol. 4660, pp. 212-221, 2002
[16] J. Konrad and P. Agniel, "Artifact reduction in lenticular multiscopic 3-D displays by means of anti-alias filtering," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 5006, pp. 336-347, Jan. 2003
[17] W. Tzschoppe, T. Brueggert, M. Klipstein, I. Relke and U. Hofmann, "Arrangement for two-or-three-dimensional display", US pat. 2006/0192908, issued Aug. 31, 2006
[18] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science,* vol. 194, pp. 283-287, 1976.
[19] J. Konrad and P. Angiel, "Subsampling models and anti-alias filters for 3-D automultiscopic displays", *Image Processing, IEEE Transactions on* , vol.15, no.1, pp. 128-140, Jan. 2006
[20] A. Boev, A. Gotchev and K. Egiazarian: "Crosstalk Measurement Methodology For Auto-Stereoscopic Screens", IEEE 3DTV Conference, Kos, Greece, May 7-9, 2007.
[21] Y. Yeh and L. Silverstein, "Limits of Fusion and Depth Judgement in Stereoscopic Colour Displays", *Human Factors*, vol 32(1), pp. 45-60, 1990
[22] J. Konrad, B. Lacotte, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," Tech. Rep. 97-01, INRS-Télécommunications, Feb. 1997
[23] S. Masnou, J. Morel, "Level lines based disocclusion," *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on* , vol., no., pp.259-263 vol.3, 4-7 Oct 1998
[24] W. Wang, L. Huo, W. Zeng, Q. Huang, W. Gao, "Depth image segmentation for improved virtual view image quality in 3-DTV," *Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on* , vol., no., pp.300-303, Nov. 28 2007-Dec. 1 2007
[25] S. Acton, D. Mukherjee, J. Havlicek, A. Bovik, "Oriented texture completion by AM-FM reaction-diffusion," *Image Processing, IEEE Transactions on* , vol.10, no.6, pp.885-896, Jun 2001
[26] R. Pajarola, M. Sainz, Y. Meng, "Dmesh: Fast Depth-Image Meshing And Warping". *Int. J. Image Graphics* 4(4): 653-681 (2004)
[27] M. Kraus, M. Strengert, "Depth-of-Field Rendering by Pyramidal Image Processing", *Computer Graphics Forum* 26 (3), 645–654, 2007
[28] M. Shinya, "Post-filtering for depth of field simulation with ray distribution buffer," In *Proceedings of Graphics Interface '94* (1994), pp. 59–66.
[29] M. Botsch, A. Hornung, M. Zwicker, L. Kobbelt, "High-quality surface splatting on today's GPUs," *Point-Based Graphics, 2005. Eurographics/IEEE VGTC Symposium Proceedings* , vol., no., pp. 17-141, 20-21 June 2005
[30] V. Podlozhnyuk, "Image Convolution with CUDA", White paper, Nvidia Corp, June 2007, available online at http://developer.download.nvidia.com/compute/cuda/sdk/website/projects/convolutionSeparable/doc/convolutionSeparable.pdf

[P11] A. Boev, A. Gotchev, K. Egiazarian, "Crosstalk Measurement Methodology for Auto-Stereoscopic Screens," Proc. *3DTV Conference, 2007*, pp.1-4, 7-9 May 2007 doi: 10.1109/3DTV.2007.4379396

This page left blank

CROSSTALK MEASUREMENT METHODOLOGY FOR AUTO-STEREOSCOPIC SCREENS

Atanas Boev, Atanas Gotchev and Karen Egiazarian

Institute of Signal Processing, Tampere University of Technology, Tampere, Finland
firstname.lastname@tut.fi

ABSTRACT

Autostereoscopic displays utilizing slanted lenticular sheets produce specific artifacts. These artifacts affect the perception of a 3D scene, and are caused by a process which can be modeled as inter-channel crosstalk. We propose methodology for measuring such a crosstalk for arbitrary multiview 3D display. The measured data might be used for optimizing multiview image sets for a given display.

Index Terms— 3DTV, multiview, auto-stereoscopic displays, inter-view crosstalk, crosstalk measurement, slanted lenticular sheet

1. INTRODUCTION

Stereoscopic 3D-perception is possible when each eye of the observer sees the scene from a slightly different perspective. There are various approaches to replicate this effect on a raster screen, in order to create the illusion of a real 3D scene being displayed [1], [2], [3].

Three-dimensional displays which create 3D effect without requiring the observer to wear special glasses are called autostereoscopic displays. The most popular ones, so called multiview 3D displays, work by simultaneously showing a set of images ("views"), each one seen from a particular viewing angle along the horizontal direction (Fig 1a) [4]. Such effect is achieved by adding an optical filter, which alters the propagation direction for the information displayed on the screen. A number of techniques exist – parallax barriers, spherical and lenticular lenses, the latter being the most common one [1]. Depending on the design parameters, various tradeoffs between screen resolution, number of views and optimal observation distance exist [1], [2], [3].

From each particular direction, only a part of the screen subpixels is seen (Fig 1b) – the one that contributes to the corresponding view. This way, it is possible each eye of the observer to see different picture, in order 3D illusion to be created. Each view is seen from a number of observation positions, and the whole set of views is consecutively repeated along the horizontal observation axis. The transition between the two outmost views produces several zones of observation, where double images are seen [4], producing characteristic "jump" in the next set of views. However, we noticed that on

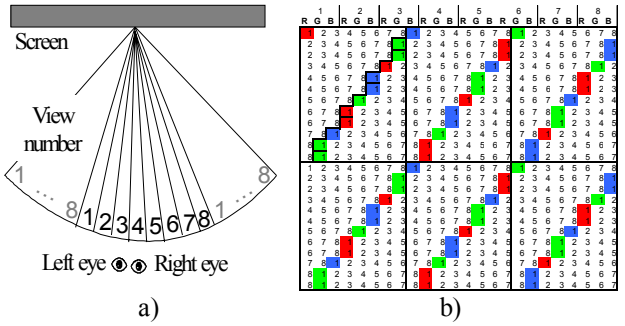


Figure 1. a) Views on an autostereoscopic screen; b) example set of subpixels corresponding to a particular view.

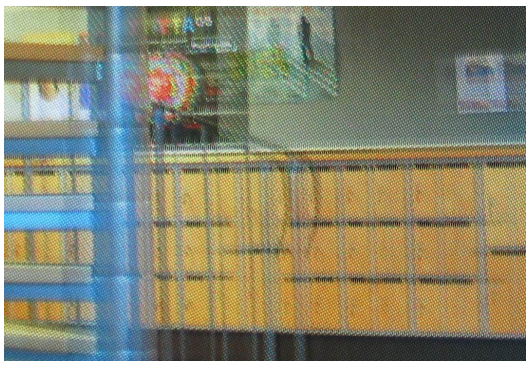
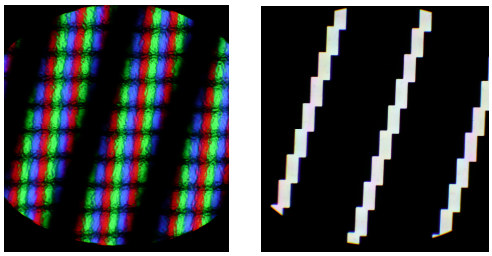
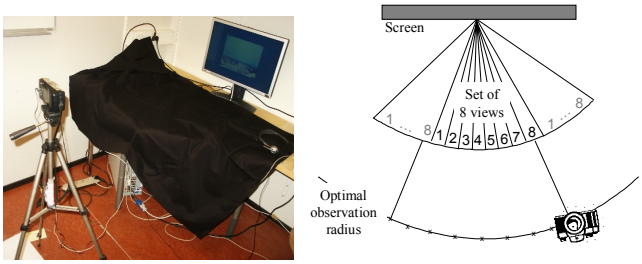


Figure 2. Snapshot of "double edges" artifacts on X3D-23" display. Snapshot is taken from observation point close to the center of view number 4.



a) b)

Figure 3. Micrograph photo of a) subpixels under the lenticular lens and b) rectifying mask



a) b)

Figure 4. a) Measurement setup; b) observation points

some scenes double images are seen from almost any observation angle.

2. PROBLEM STATEMENT

In order to balance the horizontal versus vertical resolution of an autostereoscopic displays, a slanted lens array is used [5]. This causes the subpixels of a view to appear on non-rectangular grid as illustrated in Fig. 1b. Specially designed filters need to be used to prevent the aliasing caused by subsampling on such a grid [6], [7]. Despite antialiasing, we observed another effect to be much more pronounced – parts of the neighboring views are always seen in the current view. Depending on the scene, this might produce irritating "double edges" artifacts, which destroy the 3D perception, as seen in Fig. 2.

Due to the slant of the lenticular lenses, the lens elements can not cover the subpixels' boundaries exactly. A micrograph photo (Fig. 3a) shows that some subpixels appear only partly in the current view. Furthermore, the center of the lens element is going to appear arbitrary displaced over a subpixel triplet – this causes the optimal observation point to be slightly shifted for different pixels of a certain view. Additionally, some vendors broaden the observation angle of a pixel, in order to create more uniform view [8]. All these effects cause parts of subpixels that belong to other views to be cast towards the current view. Additionally, this might introduce coloring artifacts, but this effect is barely visible on a micrograph photo, as in Fig. 3b, and not at all on a large scale (cf. Fig. 6).

The contribution of other views' subpixels depends on many production parameters, such as the design of the lenticular sheet, distance between the sheet and the pixels, and precise placement of the sheet over the screen. Such parameters are rarely available to the screen users, and might be unavailable even to the vendor – for example some companies sell lenticular lens sheets separately. Furthermore, the contribution of a single pixel is difficult to be measured, especially from the optimal observation distance. It is easier to model the process on a larger scale as crosstalk, similarly to the approach in [3].

Any algorithm that aims to mitigate crosstalk artifacts would need knowledge of the characteristics and amount it. Thus, there is a need of a measurement methodology which would assess the crosstalk between the views of an arbitrary multiview 3D screen.

3. CROSSTALK MEASUREMENT

3.1. Measurement set-up

The screen we used for the measurements was X3D-23 – 23" 8-view 3D display produced by NewSight GmbH. The screen was placed in a dark room, various test images were displayed on it, and snapshots were taken, using computer controlled camera.

3.1.1. Finding the observation points

The camera was set on the same height as the center of the screen. Only the subpixels designated in the manual as corresponding to a certain view were turned on maximal brightness, all other subpixels were turned off. With that image on the screen, there was a precise point, from which the screen was seen fully lit. This point was marked as the optimal observation point of the current view. The process was repeated for all the views, and the optimal observation points were marked on the floor, with the aid of a laser pointer attached to the tripod. The set of optimal observation points were laying, at even intervals, on an optimal observation radius, with center on the vertical axis through the middle of the screen. We designated these points as centers of the corresponding observation windows, and added three additional points between

each two view centers. Since the center of window 4 was straight in front of the screen, this resulted in a total of 33 observation points, point 17 being the center of window 4.

3.1.2. Measurement automation

The snapshots were taken by consumer camera, connected to the computer through a data cable. For automation of the acquisition process, we used free software called "PhotoPC" available on the Web [9], which operates with a wide range of photo cameras. It can set the picture acquisition parameters, perform a snapshot, and download the ready image to the computer. The focus and the aperture were fixed, and the shutter speed was experimentally set to allow good dynamic range of the photos without saturation. We prepared automated script which displays test images on the screen, takes snapshots, and optionally calibrates the results to eliminate added light. Then, the script finds the screen on the photo, compensates it for projective distortion, and performs statistical measurements. Since the acquired images were very smooth, and in order to eliminate the camera noise, local means on a 16×16 grid over the screen were measured. This resulted in 256 values per test image. The process was repeated for all test images, after which the camera was moved to the next observation point.

3.1. Measurement of crosstalk versus viewing angle

In order to measure the crosstalk, we decided to measure the individual contribution of each view towards each observation point. We prepared ten test images. Two images were used for calibration: $I_{\max}$ – with all subpixels of all views set on maximal brightness; and $I_{\min}$ – with all subpixels of all views turned off. Eight images were used for the measurement – in each one only the subpixels corresponding to a certain view set on maximal brightness ($I_1 \ldots I_8$). From each observation point we took snapshots of each test image, and computed the local means over a 16×16 grid as explained before. The output was scaled to the range [0..1] by using the formula:

$$S_{n_{x,y}} = \frac{I_{n_{x,y}} - I_{\min_{x,y}}}{I_{\max_{x,y}} - I_{\min_{x,y}}} \qquad (1)$$

The final output of the experiments was a 4-D matrix, with dimensions 16×16×8×33, containing contribution coefficients k along screen horizontal and vertical axes, view number and observation point, respectively. Slices of this matrix, showing the local means over the screen surface are shown in Fig. 6.

The presence of crosstalk is clearly visible in Fig. 7a, and it is generally similar to crosstalk measurements of another screen, presented in [3]. The figure shows that for any observation point, the main contribution comes from two neighboring views, which is also in accordance with [3]. However, Fig. 7c and Fig. 7d demonstrate that it is also needed to study the crosstalk along the screen surface. For a given observation point, the layer that contributes the most to the crosstalk, as well as the amount of it, vary along the horizontal and vertical axes of the screen. In addition, our measurements show that the X3D-23" screen has different (more pronounced) crosstalk than the screen studied in [3].

3.2. Measurement of crosstalk depending on pixel value

The previous experiment measured the contribution of each view, based on maximum pixel values. Next, we decided to measure the contribution dependence based on pixel values. As shown in Fig. 7a, the most significant contribution for certain observation point comes from the respective central view and its two closest neighbors. For example, for observation point 17, we measure the contribution of View 4 (central for this observation point), and its two neighboring views, View 3 and View 5. Such measurements give information what pixel values in the neighboring views will produce the same apparent brightness as certain pixel value in the central view.

We ran series of measurements, in which all pixels in a certain view were set to a certain value, and the pixels of another

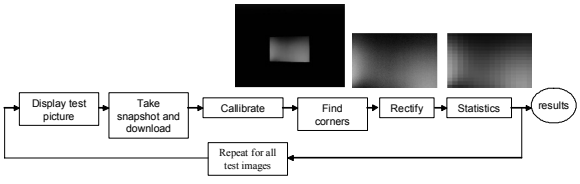


Figure 5. Measurement automation

Figure 6. Crosstalk measurements from observation point 17 – scaled local means over the screen for each view. a) S_3, b) S_4, c) S_5, d) S_6, e) S_7. The values in S_1, S_2 and S_8 are close to zero.

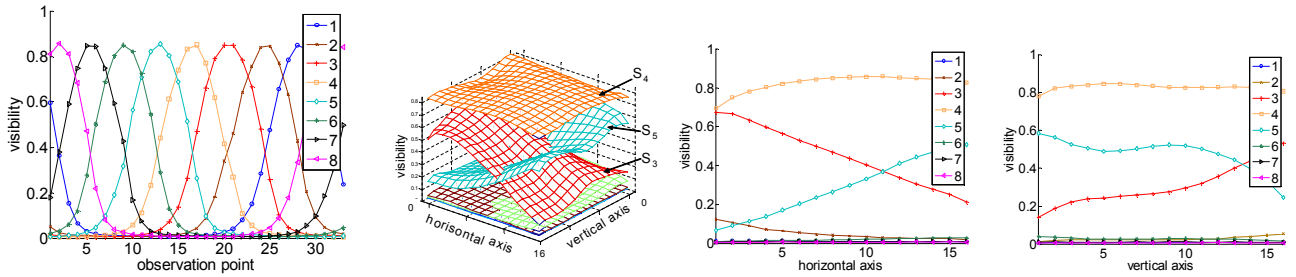


Figure 7. Measurement results - contribution of each view: a) for all observation points, measured in the center of the screen, b) towards observation point 17, c) towards observation point 17, along the horizontal axis, and d) towards observation point 17, along the vertical axis

view were gradually changed from zero to maximum. Three series of experiments were made, View 3 vs. View 4, View 4 vs. View 3 and View 5 vs. View 3. The results are shown in Fig.8 a), b) and c) respectively. Experimentally, we found that the measured value y for all measurements is closely approximated as sum of two functions – one depending on combination of input values (x_{total}) and another depending only on combination of crosstalk coefficients (k_{total}):

$$y = \frac{f_{\text{logistic}}(x_{\text{total}})}{k_{\text{total}}} + f_{\text{gauss}}(k_{\text{total}}) \qquad (2)$$

The two functions are logistic and Gaussian functions, respectively:

$$f_{\text{logistic}}(x_{\text{total}}) = 72 - \frac{72}{1 + e^{\frac{x_{\text{total}} - 201}{45}}} \qquad (3)$$

$$f_{\text{gauss}}(k_{\text{total}}) = a e^{-\frac{k_{\text{total}} - b}{c^2}} + d \qquad (4)$$

The parameters of the Gaussian function are approximated by using k_{total} :

$$a = -76k_{\text{total}} + 77.6; \quad b = 94.5k_{\text{total}}^{-0.34}; \quad c = 75; \quad d = 2.85k_{\text{total}} - 2.85 \qquad (5)$$

Finally

$$x_{\text{total}} = \sqrt{x_1^2 k_1^2 + x_2^2 k_2^2} \qquad (6)$$

$$k_{\text{total}} = \sqrt{k_1^2 + k_2^2} \qquad (7)$$

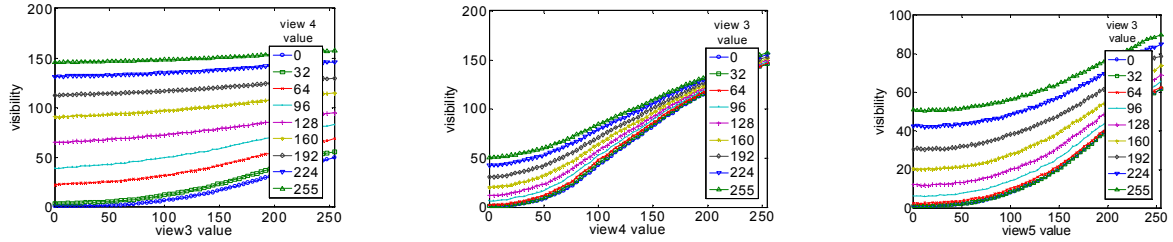


Figure 8. Measurement results – contribution of neighbours depending on pixel value: a) View 3 vs. View 4, b) View 4 vs. View 3 and c) View 5 vs. View 3. Measurements are done from observation point 17, View 4 is the central view.

where x_1, x_2 are the pixel values from the two views, and k_1, k_2 are the corresponding crosstalk coefficients for each view as measured in the experiment from Section 3.1

Figure 9. Example for scene dependant crosstalk mitigation – selective low-pass according to the inter-view differences.

4. CONCLUSIONS AND FUTURE WORK

By design, multiview 3D displays which utilize slanted lenticular sheet have a certain setback – subpixels which belong to a certain view are partly seen in the neighboring views. This effect can be modeled as inter-view crosstalk, and it introduces annoying artifacts, which are scene dependant, and may hinder proper 3D perception. The parameters of this crosstalk are rarely available to the end users, if at all. By introducing a methodology for crosstalk measurement of arbitrary multiview display, we aim at helping content creators and end users in optimizing 3D scenes for a given monitor. Future work will study the possibility of adaptive filtration of multiview image sets, in order to mitigate the effect of crosstalk. Example output of such filter is shown in Fig. 9.

ACKNOWLEDGEMENT

This work is supported by EC within FP6 (Grant 511568 with the acronym 3DTV) and by the Academy of Finland, project No. 213462 (Finnish Centre of Excellence program (2006 - 2011)).

REFERENCES

[1] L. Onural, T. Sikora, J. Ostermann, A. Smolic, M. R. Civanlar and J. Watson: “An Assessment of 3DTV Technologies,” *NAB Broadcast Engineering Conference Proceedings 2006*, pp. 456-467, Las Vegas, USA, April 2006.

[2] P. Surman, I. Sexton, R. Bates, W. K. Lee, K. Hopf, and T. Koukoulas: “Latest Developments in a Multi-User 3D Display,” in *Proc. SPIE Vol. 6016, Three-Dimensional TV, Video, and Display IV* , 2005.

[3] R. Braspenning, E. Brouwer and G. de Haan, “Visual quality assessment of lenticular based 3D-displays”, in *Proc of 13 European Signal Processing Conference,* EUSIPCO 2005, Turkey, September 2006

[4] X3D-23” Users’ Manual. NewSight GmbH. Firmensitz Carl-Pulfrich-Str. 1 07745 Jena, 2006

[5] C. van Berkel, “Image preparation for 3D-LCD,” in *Proceedings of SPIE*, 1999, vol. 3639, pp.84-91

[6] J. Konrad and P. Agniel, "Artifact reduction in lenticular multiscopic 3-D displays by means of anti-alias filtering," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 5006A, pp. 336-347, Jan. 2003

[7] J. Konrad and P. Agniel, "Non-orthogonal sub-sampling and anti-alias filtering for multiscopic 3-D displays," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 5291, pp. 105-116, Jan. 2004

[8] W. IJzerman et al., “Design of 2d/3d switchable displays,” in *Proc of the SID, v*olume 36, Issue 1, pp. 98-101, May 2005

[9] http://photopc.sourceforge.net/

This page left blank

[P12] A. Boev, M. Georgiev, A. Gotchev, N. Daskalov and K. Egiazarian "Optimized visualization of stereo images on an OMAP platform with integrated parallax barrier auto-stereoscopic display", in *Proc. 17th European Signal Conference EUSIPCO 2009*, Glasgow, Scotland, August 20 09

This page left blank

OPTIMIZED VISUALIZATION OF STEREO IMAGES ON AN OMAP PLATFORM WITH INTEGRATED PARALLAX BARRIER AUTO-STEREOSCOPIC DISPLAY

Atanas Boev[1], *Mihail Georgiev*[1], *Atanas Gotchev*[1], *Nikolay Daskalov*[2], *Karen Egiazarian*[1]

[1]Department of Signal Processing, Tampere University of Technology
P. O. Box 553, FI-33101, Tampere, Finland
phone: + 358 3 3115 4349, fax: + 358 3 3115 3857, email: firstname.lastname@tut.fi
[2]MM Solutions Ltd.
Izgrev, 15 Tintiava Str, Sofia 1113, Bulgaria
phone: +359 2 868 8162, fax: +359 2 962 4404, email: ndaskalov@mm-sol.fi

ABSTRACT

In this paper, we describe a system for optimized visualization of stereo images on a mobile platform. The system utilizes a front camera, and face and eye tracking to find the position of the observer's eyes. Depending on this position, the left and right views targeting the corresponding eyes are maintained properly based on measured optical characteristics of the used parallax-barrier 3D display.

An efficient implementation on the OMAP 3430 platform is targeted by splitting the processes of face and eye detection between the ARM and DSP cores.

The final system allows for dynamic switching the display between 2D and 3D mode and swapping the left and right views so to avoid high cross-talk between view channels and reverse stereo effect.

1. INTRODUCTION

Stereoscopic video content has became more and more popular and available in the form of 3D movies for 3D movie theatres [1], [2], and through 3D display solutions for home [3], [4] and mobile entertainment [5], [6], [7].

The mobile use of 3D video content is especially challenging since it requires creating an immersive 3D effect on a small display and processing big amount of data in a power-constrained handheld device. Autostereoscopic displays requiring no special glasses to deliver the 3D effect have been considered attractive for mobile 3D devices. Such displays however, suffer from 3D artefacts usually related with the position and angle the display is observed from. Special image processing methods are needed to prepare the images for such displays and to mitigate the corresponding artefacts. In this paper, we propose a system, which optimizes the 3D imaging to adapt it to the observation angle of the user. The system is based on OMAP 3430 and employs front-facing camera and face and eye-detection, to find the user's position and angle with respect to the screen in order to create the 3D image accordingly. The paper is organized as follows: First, we briefly present the mobile auto-stereoscopic displays, the respective artefacts and the particular display we deal with. Then, we present the suggested system for optimized visualization. Section 4 presents the implementation details concerning the face and eye detection module on the OMAP platform.

2. MOBILE 3D DISPLAYS

There are two important requirements for mobile 3D display: to create 3D effect without the need of special glasses, and to be able to switch back to "2D mode" when 3D content is not available. Autostereoscopic displays create glasses-free 3D effect by emitting different images towards each eye of the observer. In such displays, a standard portable TFT display is used to generate the images and an additional optical filter is used to redirect the light from the pixels. Thus, groups of pixels (denoted as *view*) are seen only from a specific angle. For mobile devices, normally watched by single observer, two independent views are sufficient for satisfactory 3D perception. In order to be shown on a stereoscopic display, the images intended for each eye should be spatially multiplexed. This process is known as *interleaving* [8], and depends on the parameters of the optical filter.

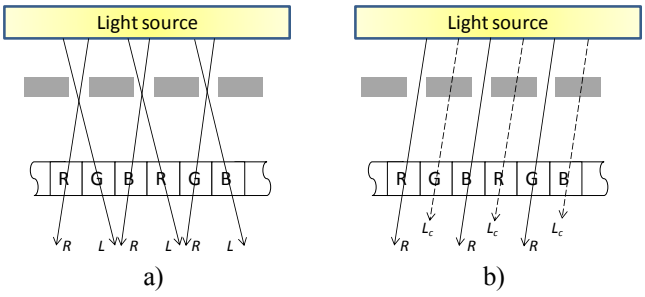


Figure 1. Parallax barrier: a) operation principle and b) crosstalk towards the optimal observation angle.

The most common approach for separating the images intended for each eye utilizes a layer called *parallax barrier*. The barrier blocks the light in certain directions as shown in Figure 1a. Two separate views are formed as a result. Due to the repetitive structure of the parallax barrier, each view is seen from a number of observation angles, as illustrated in Figure 2. In order to perceive proper 3D image, the observer should be properly positioned with respect to the display (e.g. positions "1" or "2" but not "3" in Figure 2).

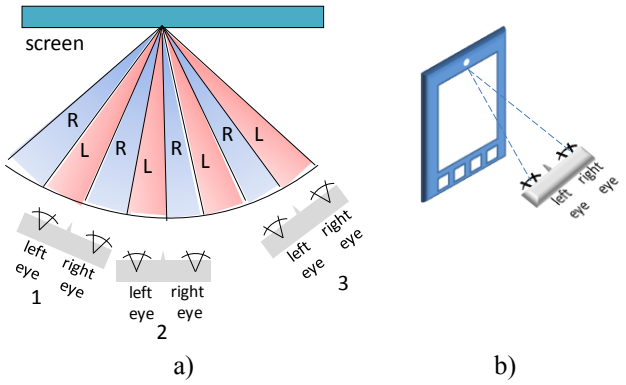


Figure 2. Position of the observer in respect to the display: a) visibility zones of views and b) position of the user as detected by the front facing camera

The parallax barrier is a cheap technology providing 2D backward compatibility through switching off the barrier.

2.1 Visual artefacts in parallax barrier-based displays

Crosstalk

Crosstalk is the effect of mixing the views. It is caused by imperfect optical separation of the views. The visual manifestation of crosstalk is a double-contoured, "ghost" images which significantly reduce the perceived 3D quality. There are two causes of crosstalk in parallax barrier-based displays. *First*, it arises when the display is observed from a position between two observation zones. The visibility of each view gradually changes as a function of the observation angle, as exemplified in Figure 3. At a certain angle, pixels of one view are fully visible, while the pixels of the other are fully covered by the parallax barrier. Such *optimal observation angle* for view 1 is marked by "I" in Figure 3, while the optimal observation angle of view "2" is marked by "III" in the same figure. At angle "II", both views are only partially covered by the barrier causing what we call *inter-zone crosstalk*. It reaches minimum at the optimal observation angle of a view, and maximum on the bisection between two neighbouring optimal observation angles. With respect to the inter-zone crosstalk, there are "high quality" areas, with no noticeable crosstalk (areas "A" and "C" in Figure 3b) and "low quality" areas ("B" in Figure 3b), where crosstalk prevents from proper 3D perception.

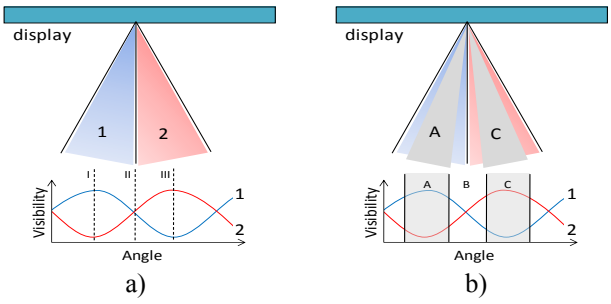


Figure 3. Crosstalk versus observation angle: a) visibility of a view as a function of the observation angle and b) "high-quality" zones with low inter-zone crosstalk

The second cause for crosstalk is the transparency of the parallax barrier. It is usually implemented as a second, not fully opaque, LCD layer. Even at an optimal angle, part of the light passes through the barrier as illustrated in Figure 1b. This amount of *minimum crosstalk* is always presented in view (cf. Figure 3).

Pseudoscopy

Regarding Figure 2a, positions "1" and "2" are proper for perceiving 3D effect. However, at position "3" in the same figure, an observer will see the "left" image with the right eye and vice versa, thus perceiving a *pseudoscopic image* (aka *reverse stereo*). Both the observation zones of the two views and the correct and pseudoscopic positions alternate. In between each correct or pseudoscopic zone an inter-zone crosstalk is perceived as exemplified in Figure 4. Moving away from a correct observation (e.g."C" in Figure 4), the observer passes through a zone where high crosstalk is visible, and then falls into zone with low amount of crosstalk, but incorrect (pseudo) stereo ("P" zone). This effect causes what is perhaps the biggest inconvenience with parallax barrier-based displays. Most observers have the instinctive ability to move away from zones where low-quality, ghosting impaired image is seen. However, pseudoscopic stereo is not immediately perceived as "bad", which might cause a user to stay at an angle maintaining low crosstalk but wrong stereo.

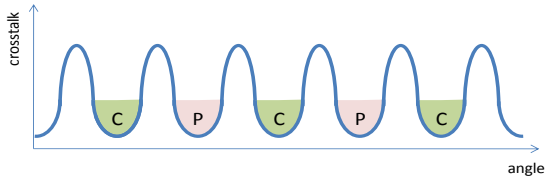


Figure 4. "Correct stereo" and "Pseudoscopic stereo" zones as function of the observation angle

2.2 Stereoscopic 3D LCD in our system

Technical data

The display model used in our system is Stereoscopic 3D LCD MB403M0117135, produced by masterImage [7]. It is 4.3" WVGA autostereoscopic display with switchable parallax barrier, which can operate in 2D or 3D mode. Additionally, the parallax barrier of the 3D LCD module can be switched between "3D horizontal" and "3D vertical" mode, thus operating in landscape 3D or portrait 3D mode. Two signals, "Chip Select" and "Mode Select" determine the mode of the display. The combinations are given in **Table 1**.

Table 1 – display modes of 3D LCD module

CS (Chip Select)	MS (Mode Select)	Display mode
Low	Low	3D Horizontal
Low	High	3D Vertical
High	Low	2D
High	High	2D

Artefacts quantified

We have measured the crosstalk and the angles between the observation zones using the methodology in [9], [10]. We found the optimal observation distance of the display to be 42 cm. At the optimal observation distance, the optimal observation points of each view found are shown in Figure 5.

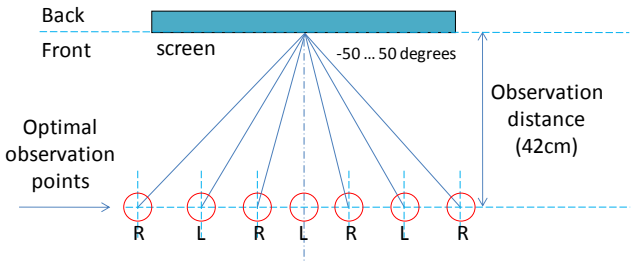


Figure 5. Observation points used in measurements

The angle between two neighbouring optimal observation vectors is 9.4°. The minimum crosstalk is 9%, and it is symmetrical with respect to the channels. The results of the crosstalk measurements at the optimal observation points of both views are shown in Figure 6.

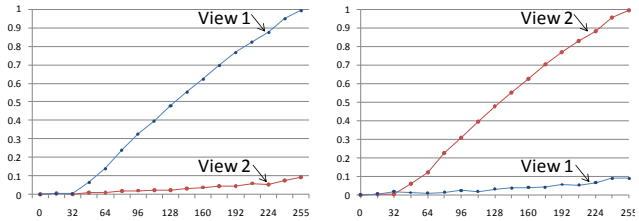


Figure 6. Minimum crosstalk: a) view 2 introduced in view 1 and b) view 1 introduced in view 2

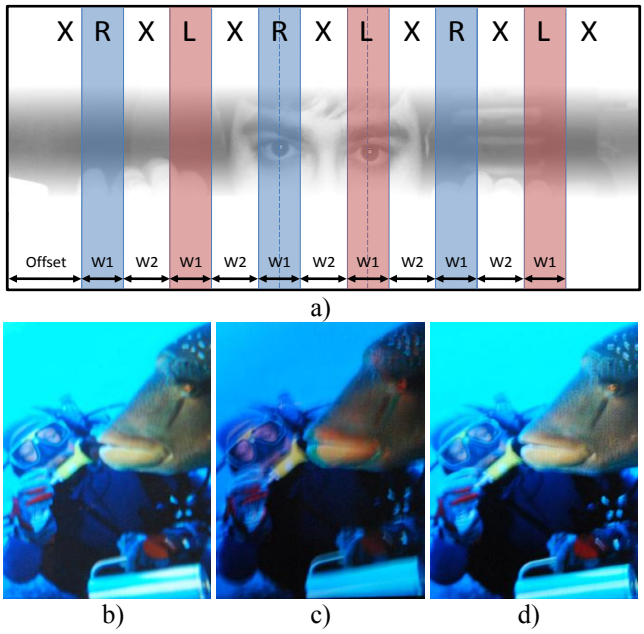


Figure 7. a) Map of "high quality" observation zones with low inter-zone crosstalk, b) image seen from area "L", c) image seen from area "X" and d) image seen from area "R" of the map

By subjective evaluation, we measured the width of the "high quality", ghost artefact-free areas ("A" and "C in Figure 2b). A test 3D image was shown on the display, and a front facing camera was mounted on the device as shown in Figure 2b. An observer assessed the image looking with one eye from the optimal observation point of one of the views. Then he started moving to the left, till noticeable ghosting was appeared in the image. The position of the pupil was recorded by taking a snapshot of the observer's eye with the camera. The right border of the ghost-free was found in a similar manner. The process was repeated for all optimal observation points, which resulted in a map of areas where the pupil of the observer must reside in order to perceive image with no hosting artefacts. The measured map is shown in Figure 7a. When the camera operates in VGA resolution, the width of ghost-free zones (marked with "W1" in Figure 7) is 20px, separated by crosstalk-impaired zones (marked with "W2") of 31px each. Images in Figure 7b, c and d show photos taken from zones "L", "X", and "R" correspondingly, and give example of the inter-zone crosstalk observed in between the "high quality" zones.

3. VISUAL OPTIMIZATION FOR 3D LCD DISPLAY

We propose a system for visual optimization of stereo imagery for an autostereoscopic display. The system is based on OMAP 3430 SDP and integrates a 3D LCD module and front-mounted camera with VGA resolution. The system tracks the position of the observer's eyes, and adapts the system to avoid three cases of visual discomfort.

Reverse stereo is avoided by simply flipping the left and right channel based on eye detected being at the opposite view zone, see Figure 8a. The pseudoscopic regions (marked with "P" in Figure 4) are replaced with zones where both channels are flipped (marked with "F" in Figure 8a), thus allowing correct stereo image to be perceived.

Ghosting artefacts are avoided by switching the display into "2D" mode, in cases when the observer's eyes falls into area with pronounced crosstalk, where the 3D perception is anyway impossible (see in Figure 7a, areas marked by 'X' and Figure 8b).

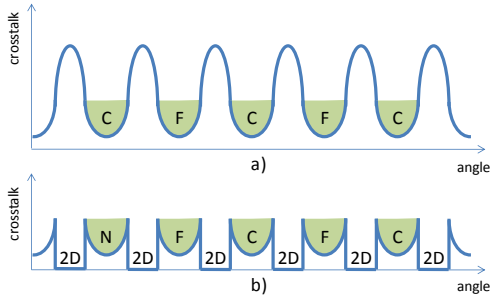


Figure 8. 3D image correction following the position of the observer's eyes: a) correction for the pseudoscopic regions, and b) correction for the regions with inter-zone crosstalk

Finally, making use of the 3D LCD module ability to switch between horizontal 3D and vertical 3D modes, our system selects 3D mode and scene orientation according to the orientation of the observer's eyes, as illustrated in Figure 9. When the face of the observer is not in horizontal of verti-

cal direction in respect to the display, 3D effect is not possible, and thus the system switches the display into 2D mode.

The block diagram of the algorithm is shown in Figure 10. It goes through the following stages:

1. Face detection is attempted four times, each time rotating the camera image at a right angle. If face is not detected, it is possible that the face of the observer is at a wrong angle or too far away from the centre of the display. In both cases 3D perception is not possible, and the system switches the display into 2D mode.

2. If face detection is successful, its direction is stored, and eye tracking is performed according to the direction.

3. The position of the eyes is matched against the map of "high quality" observation regions. The map in use is selected to match the direction of the face.

4. If both eyes are found in the corresponding regions, the system switches into 3D mode. If both eyes appear in the regions of the opposite view, the system flips the channels and switched into 3D mode. If both eyes fall into the observation zone of the same view, or at least one eye falls in an inter-zone crosstalk area, the system switches into 2D mode.

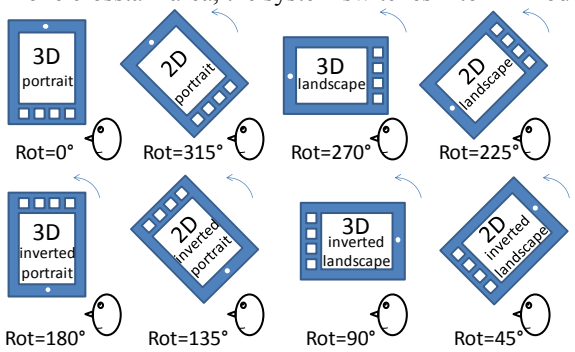


Figure 9. Selection of 3D mode and scene orientation according to the orientation of the observer's head

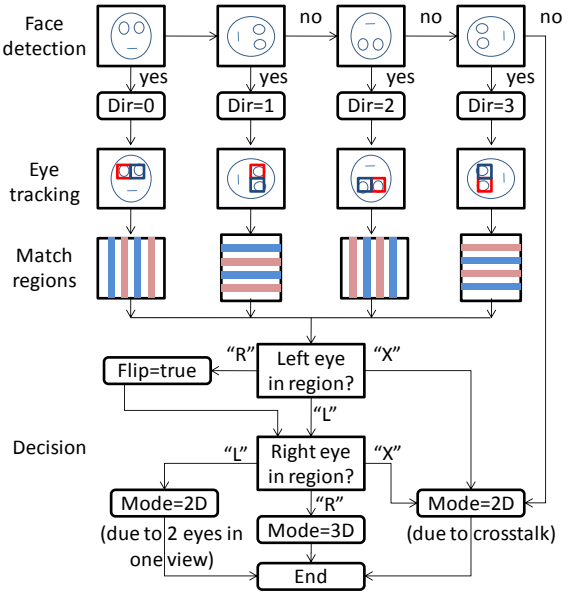


Figure 10. Block diagram of the proposed algorithm for visual optimization

4. IMPLEMENTATION

The system is implemented on the OMAP 3430 SDP running Linux OS (L12.20 baseline release). A parallax-barrier auto-stereoscopic display has been integrated to the platform [11]. OMAP 3430 is a dual-core processor, which includes general purpose ARM core and a TMS320 compatible DSP core. The ARM side provides access to C compiler and Linux environment, which allows code from existing open source libraries to be reused. The ARM side has been used for code prototyping while time-critical functions has been ported to the DSP in a block-by-block fashion. The dual-core architecture allows the output of both implementations to be compared and simplifies the debugging process.

The application processing modules have been distributed between the ARM and DSP processors as shown in Figure 11, aiming at an efficient implementation. The ARM side is engaged by the Linux OS. It is also responsible for maintaining the camera images, detecting the face and generating the stereo views. The DSP is engaged by the computationally-intensive eye detection algorithm. An effective inter-processor communication protocol is used through a queued mailbox-interrupt mechanism [16].

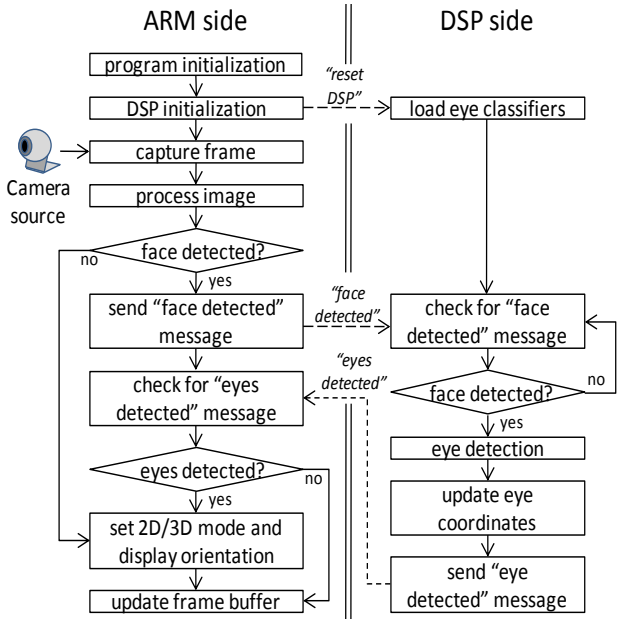


Figure 11. Application flow diagram of DSP and ARM.

4.1 Face and eye detection

We have ported an OpenCV realization of face detection algorithm by Viola and Jones [17] by modifying the classifiers to use fixed point arithmetic. Our own face detection algorithm is being ported to the OMAP as well. It is based on a two-stage hybrid technique, combing skin detection with feature-based face detection [12], [13]. In our face detection implementation, the search for faces is done for a subset of face sizes – limited by the expected facial size of an observer within the visual comfort zone for the 3D display. It applies a large-to-small scale search strategy, and the search is satisfied

by the first face found, thus ensuring that the display mode is set appropriately for the *closest* observer.

The eye detection is implemented on the DSP core. It detects the two pupils by a Bayesian classifier working on Dual-Tree Complex Wavelet Transform (DT-CWT) features. The DT-CWT has been chosen as a low-cost alternative to Gabor transform for real-time feature extraction implementation [14], [15]. The DT-CWT features are formed by a four-scale DT-CWT applied on a spatial area of 16x16 pixels around a landmark., with six differently-oriented sub-bands per scale. The resulting twenty four matrices as shown in Fig. 5 form a landmark jet [14], [15].

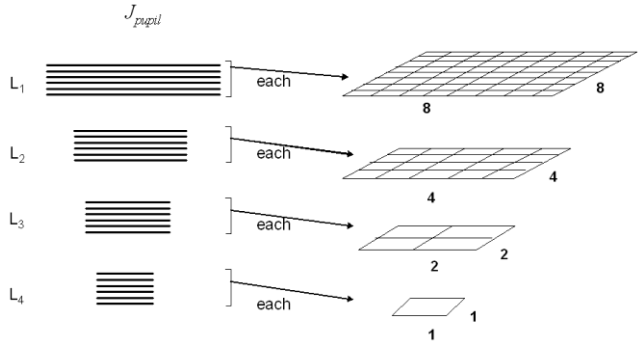


Figure 12. DT-CWT based feature jet.

Two landmark classes are modeled: pupil and non-pupil respectively. For modeling a particular landmark class, we have trained Gaussian mixture model (GMM) for each sub-band in the jet, thus leading to 24 models. We have used utmost 5 Gaussian components for each slice.

The positions of the detected eyes are returned back to the ARM by the by means of the shared SDRAM. Based on the position of the eyes, the stereo images are properly manipulated, as described in the algorithm of Section 3. The view rendering is implemented by a direct DMA operation [11].

5. CONCLUSIONS

We presented a system for optimizing the image on a 3D LCD display module by using eye-tracking and adapting to the position of the observer's eyes.

Instead of presenting an improper, low quality stereo-pair, our system switches to 3D mode only if the 3D effect is guaranteed. The resulting system will deliver 3D image only when looked at from a set of observation angles, but will avoid confusing the user by showing him low-quality image when 3D perception would not be otherwise possible.

An efficient implementation on the OMAP 3430 platform has been targeted by splitting the processes between the ARM and DSP cores. The system is part of bigger system for playing stereo video content on 3D mobile device.

6. ACKNOWLEDGMENT

This work is supported by the European Commission within the ICT Programme of FP7 under Grant 216503 with the acronym MOBILE3DTV.

REFERENCES

[1] B. Schiffman, "Movie Industry Doubles Down on 3-D", Wired magazine, April 2008, available at http://www.wired.com/techbiz/media/news/2008/04/3d_movies

[2] The Illustrated 3D Movie List, available online at http://www.3dmovielist.com/list.html

[3] "Mitsubishi Digital Electronics America Showcases Large-Screen 3D-Ready HDTV at CES 2008", press release by Mitsubishi Electric, Dec 2008, available at: www.mitsubishi-tv.com/pressreleases.html

[4] 3D DLP HDTV white paper, Texas Instruments, 2009, available online at http://www.dlp.com/hdtv/3-d_dlp_hdtv.aspx

[5] G. J. Woodgate, J. Harrold, "Autostereoscopic display technology for mobile 3DTV applications", in Proc. SPIE Vol.6490A-19, 2007

[6] S.Uehara, T.Hiroya, H. Kusanagi; K. Shigemura, H.Asada, "1-inch diagonal transflective 2D and 3D LCD with HDDP arrangement", in Proc. SPIE-IS&T Electronic Imaging 2008, Stereoscopic Displays and Applications XIX, Vol. 6803, San Jose, USA, January 2008

[7] "Stereoscopic 3D LCD Display module", Product Brochure, masterImage, 2009, available online at http://www.masterimage.co.kr/new_eng/product/module.htm

[8] Pastoor, "3D displays", in (Schreer, Kauff, Sikora, edts.) 3D Video Communication, Wiley, 2005.

[9] A. Boev, A. Gotchev and K. Egiazarian, "Crosstalk measurement methodology for auto-stereoscopic screens", Proc. of 3DTV Con, Kos, Greece, 2007

[10] A. Boev, A. Gotchev and K. Egiazarian "Stereoscopic Artifacts on portable auto-stereoscopic displays: what matters", In Proc. of VPQM 2009, Scottsdale, AZ, USA.

[11] A. Gotchev, A. Tikanmäki, A. Boev, K. Egiazarian, I. Pushkarov, N. Daskalov, 'Mobile 3DTV technology demonstrator based on OMAP 3430', submitted to DSP 2009 conference.

[12] A. Boev, M. Georgiev, A. Gotchev and K. Egiazarian, 'Optimized single viewer mode of multiview autostereoscopic display', in Proc. EUSIPCO 2008, August, Lausanne, Switzerland.

[13] V. Uzunov, A. Gotchev, K. Egiazarian, J.Astola "Face Detection by Optimal Atomic Decomposition", Proceedings of the SPIE, Volume 5916, pp. 160-171 (2005).

[14] N G Kingsbury, Complex wavelets for shift invariant analysis and filtering of signals, *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no 3, May 2001, pp. 234-253.

[15] H. Essaky Sankaran, A. Gotchev, K. Egiazarian, and J. Astola, "Complex wavelets versus Gabor wavelets for facial feature extraction: a comparative study", Proceedings of SPIE, vol. 5672, pp. 407-415, 2005.

[16] OMAP35x Applications Processor Interprocessor Communication (IPC) Module, Technical Reference Manual, TI.

[17] P. Viola, M. Jones, "Robust Real-time Object Detection," In proc. J. Of computer vision, 2001

This page left blank

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-2892-7
ISSN 1459-2045